Ali Neissi Shooshtari

Optimizing handover performance in LTE networks containing relays

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 29.4.2011

Thesis supervisor:

Prof. Jyri Hämäläinen

Thesis instructor:

D.Sc. Riikka Susitaival

Aalto University
School of
Electrical
Engineering

ABSTRACT OF THE MASTER'S THESIS

Author: Ali Neissi Shooshtari

Title: Optimizing handover performance in LTE networks containing relays

Date: 29.4.2011 Language: English Number of pages:9+79

School of Electrical Engineering

Department of Communications and Networking

Professorship: Communication Engineering Code: S-72

Supervisor: Prof. Jyri Hämäläinen

Instructor: D.Sc. Riikka Susitaival

The purpose of relays in Long Term Evolution (LTE) networks is to provide coverage extension and higher bitrates for cell edge users. Similar to any other new nodes in the network, relays bring new challenges. One of these challenges concerns mobility. More specifically, back and forth data transmission between the Donor Evolved NodeB (DeNB) and the Relay (RN) during the handover can occur.

For the services that are sensitive to packet loss, receiving all the packets at the destination is crucial. In cellular networks when the User Equipment (UE) detaches from the old cell and attaches to the new cell, it faces a short disruption. In the disruption time when the UE is not connected to anywhere, packets can be easily lost. To avoid the consequences of these packet losses, the data forwarding concept was developed and the lost packets in handover were identified and forwarded to the destination. The forwarded packets would be transmitted to the UE as it becomes attached to the new cell.

In the networks using the relays all the packets should still transfer via the DeNB. If the UE is connected to the RN and is handed over to a new cell, the unacknowledged packets between the RN and the UE which are still in the RN buffer should be transmitted back to the DeNB and onwards to the target afterwards. Furthermore, the ongoing packets in S1 interface are transmitted through the old path until the path switch occurs. This data transfer from the DeNB to the RN and again back to the DeNB increases the latency and occupies the resources in the Un interface.

In this thesis work the problem of back and forth forwarding is studied. Different solutions to overcome this challenge are proposed and simulations are performed to evaluate the proposals. The evaluated approaches showed up to considerable performance enhancement compared to the previous solutions.

Keywords: LTE, Handover, Relay, Data forwarding

Acknowledgement

This master's thesis was performed in Nomadic Lab, Ericsson research Finland.

First and foremost I would like to immensely thank Riikka Susitaival my instructor at Ericsson for her patience, impressive ideas, way of explanation and remarkable attention. I would also like to extend my deepest gratitude towards Jyri Hämäläinen my supervisor in Aalto university for his excellent supervision, technical discussions and valuable comments. It was my good fortune to work with them.

Johan Torsner my manager at Ericsson is indeed a person I am highly obliged to. It is my honour to appreciate Johan enormously not only because of giving me the opportunity to work on this astonishing project in his brilliant team, but also because of all his supports during my stay at Nomadic lab. His unique behaviour will be always in my memory and I will be always grateful for having the chance to work with him.

It is my pleasure to thank Vesa Virkki, Edgar Ramos, Stefan Wager, Kimmo Hiltunen, Mats Sågfors, Janne Peisa, Anzil Abdul Rasheed, Jose Luis Pradas and Anna Larmo, my colleagues at wireless access network lab and all my other colleagues at Ericsson research for all devoted time and memorable moments that I spent with them.

Furthermore I appreciate William Martin, the faculty's language support specialist for his humble contribution in proofreading the final version of the text and Jenni Tulensalo my study advisor at Aalto university for all her ever present guidance.

Finally, with the greatest regard I would like to express my special thank towards my family for their overwhelming love and encouragement. Hereby I dedicate this work to them.

Otaniemi, 29.04.2011

Ali Neissi Shooshtari

Contents

Abbreviations

3GPP	3rd Generation Partnership Project
ACK	Acknowledgement
AF	Amplify and Forward
AM	Acknowledged Mode
ARQ	Automatic Repeat Request
AS	Access Stratum
BS	Base Station
BSR	Buffer Status Report
CCCH	Common Control Channel
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CN	Core Network
C-RNTI	Cell Radio Network Temporary Identifier
DF	Decode and Forward
DL	Downlink
DRB	Data Radio Bearer
DRX	Discontinuous Reception
EDGE	Enhanced Data rates for GSM Evolution
eNB	Evolved NodeB
EPC	Evolved Packet Core
EPS	Evolved Packet System
ETWS	Earthquake and Tsunami Warning System
E-UTRAN	Evolved-UTRAN
FDD	Frequency Division Duplex
FMS	First Missing SDU
FTP	File Transfer Protocol
GERAN	GSM EDGE Radio Access Network
GPRS	General Packet Radio Service
GTP	GPRS Tunneling Protocol
GUTI	Globally Unique Temporary Identity
GW	Gateway
HARQ	Hybrid Automatic Repeat reQuest
HFN	Hyper Frame Number
HO	Handover
HSPA+	High Speed Packet Access Evolution
HSS	Home Subscription Server
IMS	IP Multimedia Sub-system
IP	Internet Protocol
LOS	Line of Site
LSB	Least Significant Bit
LTE	Long Term Evolution
MAC	Medium Access Control

MAC-I	Message Authentication Code for Integrity
MAG	Mobile Access Gateway
MBSFN	Multimedia Broadcast Single Frequency Network
MIB	Maste Information Blocks
MIMO	Multi-Input Multi-Output
MM	Mobility Management
MME	Mobility Management Entity
MMS	Multimedia Messaging Service
MN	Mobile Node
MSB	Most Significant Bit
MT	Mobile Terminal
MTBH	Mean Time between Handover
NACK	Negative Acknowledgement
NAS	Non Access Stratum
OFDMA	Orthogonal Frequency Division Multiple Access
PCC	Policy and Charging Control
PCEF	Policy Control Enforcement Function
PCRF	Policy and Charging Resource Function
PDCP	Packet Data Convergence Protocol
PDN	Packet Data Network
PDSCH	Physical Downlink Shared Channel
PDU	Protocol Data Unit
P-GW	PDN Gateway
PSDN	Public Switched Data Network
QoS	Quality-of-Service
RACH	Random Access Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Radio Bearer
RLC	Radio Link Control
RLF	Radio Link Failure
RN	Relay
RNL	Radio Network Layer
ROHC	Robust Header Compression
RRC	Radio Resource Control
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RTP	Real-time Transport Protocol
SAE	System Architecture Evolution
SAP	Service Access Point
SDU	Service Data Unit
S-GW	Serving Gateway
SIB	System Information Blocks
SINR	Signal-to-Interference plus Noise Ratio

SIP	Session Initiation Protocol
SM-R	Start Marker for Relay
SN	Sequence Number
SRB	Signaling Radio Bearer
TA	Tracking Area
TB	Transport Block
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TE	Terminal Equipment
TM	Transparent Mode
TTT	Time to Trigger
UDP	User Datagram Protocol
UE	User Equipment
UICC	Universal Integrated Circuit Card
UM	Unacknowledged Mode
UP	Uplink
USIM	Universal Subscriber Identity Module
UTRAN	Universal Terrestrial Radio Access Network
WiMAX	Worldwide interoperability for Microwave Access
WWW	World Wide Web

1 Introduction

1.1 Overview

Wireless technologies undoubtedly can be considered as one of the fastest ever growing technologies. This growth is indicated by the enthusiasm of consumers towards using wireless technologies. The developments in wireless communications not only have brought many business opportunities, but also have increased the quality of life for typical end-users. On the other hand, network providers as well as operators have been fully involved in improving the available systems to satisfy the eagerness of customers by providing new services and enhancing the quality of them.

To provide acceptable services that can be used everywhere, there is a need for a unity between different providers and operators. This cooperation is needed in order to make different systems compatible with each other. Therefore standardization associations have been established for adopting unique pathways. So far many players have been actively contributing to standardizing the schemes and procedures for new technologies in addition to adopting better ideas for performance enhancement of current systems. Performance improvement can be achieved by optimization or by building a new environment that again can be considered as an optimization of the current ecosystem.

The emergence of the Internet in all its various aspects has found its position as a framework for almost everything. The trend to use the Internet everywhere and for everything is introducing it as one of the basic needs in daily life. Now ubiquitous Internet can be achieved by the use of wireless communications.

Wireless technologies first gave users the joy of speaking through their mobile phones regardless of their position and their situation. In the next step it was the time for data to go mobile. Providing all the functionalities such as sending emails, downloading files, video chatting, and streaming that were available using fixed wired networks, into the mobile handsets, could make dream of a seamless Internet became reality. Mobile broadband mainly makes the commuters use their idle-time effectively. It provides users with a constant access to the Internet without having to wait until reaching somewhere to be connected.

As with any other new technologies, mobile broadband has its own challenges, even though much effort has been made so far to make this come true. However, optimization for performance improvement seems to be inevitable at any time. One of the ever-present concerns in ubiquitous services in cellular networks is the mobility aspect. The service should be maintained when the user is handed over from one cell to another cell.

Considering data transfer in handover time, data should not be delayed or should not be lost; otherwise performance will be dramatically degraded. Taking care of data in handover is especially important in non-real time services where the performance is dependant and sensitive to in-order reception of data. In non-real time services, the file sizes are relatively large and missing even a small portion of the file is not desirable.

In order to increase the absolute coverage, service coverage and data rate at the cell edge, intermediate nodes have been introduced. These intermediate nodes are designed to be placed between macro base stations and the UEs and include micro, pico, femto base stations or relays. The interesting feature about relays is to be self backhauled. Not having any wire in the backhaul makes them cheaper and easier to install for the operators. Relays are usually located at cell edges or coverage holes where the received SINR is not good. They can be used as a fast and feasible solution for temporary or permanent network extension. The use of relays in indoor scenarios where the penetration loss can degrade the signal is also a matter of interest. Putting a relay as an intermediate node outside or inside the building can amplify the signal before its reception by the user. Deploying relays, however, introduces new challenges. One of these challenges concerns mobility.

The importance of mobility for providing ubiquitous services on the one hand, and on the other hand, the role of relays as cost effective plug and play nodes in a network, give us a motivation to look into the mobility challenges that deploying relays in the network brings. The problem especially addressed here is unnecessary travel of data between an eNB and a RN in a back and forth manner during handover.

When the UE is attached to the relay and is handed over to another node in the network, some amount of data that has already been transferred to the RN from the eNB should be transmitted back to the eNB and then forwarded to the destination. This problem is called back and forth data forwarding. Increased latency and wasted backhaul resources are the undesired consequences of this problem. Therefore the objective of this simulation-based study is to find and evaluate the solutions to eliminate the occurrence of back and forth data forwarding and thus decrease latency and improve resource utilization.

1.2 Thesis structure

Chapter 2 briefly discusses the LTE networks. A proper grasp of LTE structure, the functionality of different nodes, protocol stack and the way how layer 2 protocols especially Packet Data Convergence Protocol (PDCP) and Radio Link Control (RLC) work are very important in order to understand the following chapters. Therefore these topics are explained to provide enough background for the other chapters.

Chapter 3 explains the handover definition as well as the handover procedure. Having the knowledge of the handover scheme in general with all the messages and steps is necessary for understanding the rest of the thesis. The important factors in handover design and different research areas related to handover are covered.

Chapter 4 describes relays. The thesis deals with the mobility challenges that deploying relays brings to the network. This chapter gives an understanding of the behavior of relays in networks, their value, their structure, and different types of relays.

Chapter 5 thoroughly defines the problem that was mentioned briefly in the last section. The problem is formulated in this chapter and the different proposed solutions, which have the possibility of solving the problem, are analyzed in detail.

Chapter 6 demonstrates all the findings in this thesis. A system-level simulation has been performed to evaluate all the proposals discussed in Chapter 5 using the full protocol simulator. The chapter is followed by the analysis of the results.

Chapter 7 concludes the whole work and makes recommendations for promising areas of future research.

2 LTE

2.1 Historical background

The facilities for using mobile Internet should be developed in order to satisfy the demands of people for accessing the Internet while they are on the move. Mobile broadband was introduced as an answer to this demand and very soon solutions like LTE and Worldwide Interoperability for Microwave Access (WiMAX) were considered to provide benefits for subscribers. Mobile communication has been one of the technologies that has grown very rapidly since its invention.

The 1G system was designed only for voice using the analog circuit switch networks. 1G could cover the limited regions and it did not provide global coverage; also low capacity was another limitation for this system. To solve these problems 2G systems with the usage of digital circuits were introduced. These systems were targetted at voice communications as well as text messages and they gained a lot of success due to the digital backbone, as well as the circuit and packet switch capability which supported higher data rates complying the technology improvement and global roaming.

2G was established based on voice communication and it had a very low data rate. This slowness of data data transport was considered as a bottleneck for operators who wished to provide new data services to their costumers. That was why the 2.5 G system was introduced to increase the data rate first to 56 kbps, and then up to 114 kbps. Global System for Mobile Communications (GSM) Enhanced Data Rates for Global Evolution (Edge) was also introduced in the second generation to provide higher data rates up to 236.8 kbps. These higher data rates opened up new opportunities for designing and implementing data services such as Multimedia Messaging Service (MMS) and email.

Although in 2G most of the traffic was voice and for many years operators invested in improving the capacity for voice communications, communicating with text messages became more and more popular. Transferring from analog to digital technology introduced many benefits such as using coding techniques to obtain better performance as well as increase spectral efficiency. An increase in data usage brought different types of networks such as the 3G network, with support for IP and improved data communications. After 3G was welcomed by the subscribers, network providers and operators tried to introduce new solutions by the use of new technologies for good support of data and usage of Internet everywhere. The building of 3G was started in 1985 and it took 12 years to complete. Now High Speed Packet Access Evolution (HSPA+) systems can support 42 Mbits/s on downlink and 5.8 Mbits/s in uplink [1].

In 2004, the process for 4G started with LTE and 802.16e being the two candidates introduced. Later, these two became known as 3.9G since they could not satisfy all the requirements for 4G systems. LTE Release 8 has a throughput of 300Mbps (downlink) and 75 Mbps (uplink), low latency of around 5 ms in radio access, and a 100 ms connection setup time [2]. LTE-Advanced (Release-10 version of LTE) and 802.16m as the next version of 802.16e were introduced as 4G technologies. Also LTE-Advanced uses key techniques such as carrier aggregation, enhanced DL/UL multi-antenna transmission, support for HetNet, CoMP transmission and reception, and support 3 Gbps (downlink) and 1.5 Mbps (uplink) [3] [4].

4G standardization is developing very fast. 4G is assumed to support different technologies, different terminals and to serve them in a seamless way. Seamless services introduce the challenge of handovers either intra-network handovers or inter-network handovers that should be well studied and standardized. The other issue which is dramatically important is the increase in power consumption due to the processors that are used at different levels as well as the content rich services. Usage of multi hop and cooperative communications can help to optimize the power consumption. Designing the terminals might also be a challenge for supporting multi-standards and multi-features. Due to diverse network structures, interference might also be an issue. When using Multi Input Multi Output (MIMO), the effective antenna placement is considered as a potential area that provides some gains [1].

2.2 General overview

In LTE, evolution in system architecture was a consequence of evolution in the radio interface. The evolution path of system architecture was towards supporting completely packet-switched services and to comply with some radio design goals such as hard handover instead of soft handover which existed in HSPA. Also all radio functionalities were concentrated in nodeBs to make the architecture more flat. Changing the system architecture followed different targets such as:

– Optimization of system architecture for packet-switched services, as in LTE there was no support for circuit-switched services

– Support for higher throughput comparing to former technologies due to the end-user higher bitrate demands

– Support for improvement in response time for activation and bearer set-up

– Packet delivery delay reduction

– Support for simplification of the whole system

– Interworking with other 3GPP access networks and other 3GPP technologies

As many of the targets mentioned above can be achieved by using flat architecture, it was proposed for LTE. In a flat architecture less nodes are involved and therefore the latencies decrease and performance increases. To talk about different elements in the network we divide the network into four different domains that are named as the UE domain, Evolved Universal Terrestrial Radio Access Network (E-UTRAN), Evolved Packet Core Network (EPC), and the Service domain.

Internet Protocol Connectivity Layer is represented by the three first domains: UE, E-UTRAN, and EPC, these three domains together are called the Evolved Packet System (EPS). EPS is highly optimized for providing IP-based connectivity and all the services are offered on top of IP. There are no circuit-switched nodes or interfaces in E-UTRAN and EPS at all.

E-UTRAN is comprised of only one type of node which is called the eNodeB. All the radio functionalities are concentrated to the eNodeB. E-UTRAN is organized as a mesh of eNodeBs that are connected to each other via X2 interface. EPC does not have any ciruit-switch domain and maybe this is the main architectural change. Also there is no connectivity between EPC and traditional circuit-switch networks like Public Switched Data Network (PSDN).

After this short overview in the rest of the chapter we explain LTE system performance requirements, the architecture of LTE including different logical elements in the network and their responsibilities, the protocol stack both in the control plane and the user plane, and the main functionalities of different layers. The overview will take a closer look at PDCP which is one of the sublayers of the data link layer.

2.3 LTE system performance overview

The operators should be assured that LTE has better performance compared to the former technologies and it can provide market interests. To satisfy this, certain requirements are set for LTE performance affecting the development of the technology. Here we describe these performance requirements.

2.3.1 Peak rate and peak spectral efficiency

The peak rate is defined as the maximum throughput per single user if we assume that the whole bandwidth with the highest modulation and coding scheme, as well as the maximum number of supported antennas are, allocated to a single user. In the Time Division Duplex (TDD) systems, the downlink and uplink periods peak rates are calculated separately. The maximum spectral efficiency is defined as the division of the peak rate by the used allocated spectrum. The peak spectral efficiency in downlink for a 4×4 antenna configuration is 16 b/s/Hz and for an 8×8 configuration is 30 b/s/Hz. In the uplink the peak spectral efficiency for a 2×2 antenna confiuration is 8.1 b/s/Hz and for a 4×4 antenna configuration is 16.1 b/s/Hz.

The peak data rate is a criterion that can advance one system over another one. Peak data rate, however, is not considered as a very good criteria in practice. Although it is relatively easy to design a system with high peak rates for the users near the base station, it is very hard to provide a good data rate together with coverage and mobility. Therefore there is another way for performance evaluation which is called system-level performance. A system-level simulation can be achieved using a simulator containing all the protocols, all the interfaces and several cells.

2.3.2 Cell throughput and spectral efficiency

Cell-level performance is a very important factor as it is a factor for determining the number of sites that the operator should establish. The number of sites mainly defines the cost of system deployment. The cell-level metrics are defined as the following:

- Average cell throughput [bps/cell] and spectral efficiency [bps/Hz/cell]
- Average user throughput [bps/user] and spectral efficiency [bps/Hz/user]
- Cell-edge user throughput [bps/user] and spectral efficiency [bps/Hz/user]

For the last one the metric used for the assessment is the 5-percentile user throughput of the cumulative distribution function (CDF) of the user throughput.

2.3.3 Mobility and cell ranges

User mobility with a speed of 350 km/h even up to 500 km/h is supported by LTE systems to provide a very pleasant working environment for the users who are traveling in high speed trains. This criteria emphasizes the necessity and importance of lossless handover that the service can maintain.

2.4 LTE system architecture

LTE aims to provide a seamless IP connectivity between the user and the Packet Data Network (PDN) in a network that only supports packet switched services. The mobility is fully supported by LTE and UEs should not be facing any disruption in getting services while they move. To make this happen, first an evolution in radio access and then in the core network to comply with radio access changes occurred. The evolved radio access network part is called LTE while the evolved core network part is called EPC. Besides EPC, other aspects also changed and System Architecture Evolution (SAE) was defined. EPC is a part of SAE. LTE together with SAE is known as EPS. The concept of bearers is used for routing data from a gateway in PDN to the UE. A bearer is defined as an IP packet flow with a defined quality of service which establishes between EPC and the UE.

In this section we describe the whole architecture of EPS which is divided into the core network and the access network. All the logical nodes on each side will be introduced and their functions will be defined. Then the protocol architecture in EPS will be discussed by considering the protocols in different layers both for the control plane and the user plane.

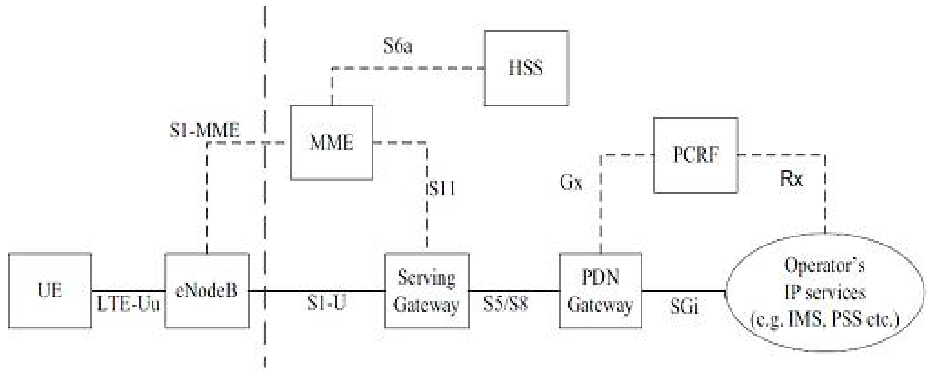

Figure 1: EPS network elements [5].

The main responsibility of EPS is to provide IP connectivity between the user and the PDN so that a user can access the Internet. As it was mentioned before, EPS uses the bearer concepts to establish the connection between the UE and PDN. The UE can use more than one bearer at the time. The bearers can have different Quality of Service (QoS) and they can be used for different purposes at the same time. For example, a UE can browse the web while it is downloading a file using FTP. The other important aspects of EPS are care of security and privacy to prevent misuses in the network. As can be seen in Figure 1, in the core network there are various nodes with different functions while in the access network there is only one type of node which is called the E-NodeB.

2.4.1 The Core Network (CN)

The CN was called EPC in former subsections. The core network is responsible for controlling the UE. It also establishes and releases the bearers. The CN has different logical nodes that are discussed in the following sections.

PDN Gateway (P-GW)

The Packet data networks gateway can be thought of as a router which is located between EPS and an external packet data network. In the sense of mobility it is considered as the highest level mobility anchor in the system and there is end-to-end IP connection between the UE and P-GW. The main responsibility of the P-GW is traffic filtering and traffic gating. P-GW should be located in a safe place in operator premises. The UE gets its IP address from the P-GW. This IP address can be used by the UE for communicating with IP hosts in external networks such as the public Internet. The IP address will be allocated to the UE whenever the UE requests for a PDN connection. Based on the need IPv4, IPv6 or both addresses can be allocated to the UE. As mentioned above, one of the responsibilities of the P-GW is to perform gating and filtering functions needed by the set of policies for the UE which is done by an element in the P-GW called the Policy Control Enforcement Function (PCEF). Charging information is collected and reported by the P-GW. In mobility, if there is a need for Serving Gateway (S-GW) change, the P-GW as the highest mobility anchor point can switch the path from the old S-GW to the new S-GW.

Serving Gateway (S-GW)

The prominent responsibility of the S-GW is switching and user plane tunnel management. In the controlling part of the S-GW has a very tiny role to play. The S-GW allocates or releases its resources based on MME or P-GW request. If the request is from the P-GW, the S-GW sends a copy of this request to the MME so that the MME can control the tunnels to the eNodeB. In handover time, the S-GW receives the command from the MME to switch the tunnel from one eNodeB to another eNodeB. MME may also command S-GW to provide tunneling resources for data forwarding from the source eNodeB to the target eNodeB, if there is a need for data forwarding. If during handover there is a need to change the S-GW, then the MME controls the tunnel change from the old S-GW to the new S-GW.

When the UE is in the connected mode, all the data flows from an eNodeB are relayed to the P-GW and vice-versa, while in the idle mode no relaying is required. If the UE is in idle mode and in the meanwhile the S-GW receives some packets from a P-GW, the S-GW starts to buffer the data and ask the MME to page the UE. After the paging and when the tunnels establish, the S-GW can transmit the buffered data. The S-GW also has the ability to monitor the data in the tunnels for charging purposes or for delivering them to authorities in case they need the data.

Mobility Management Entity (MME)

In EPC, the main control element is the MME which acts as a server and it should be located in a secure location. The MME is only operating in the control plane and it is not involved in the user plane operation at all. There is a direct logical connection between the MME and the UE in the control plane which is considered as the primary control plane connection between the UE and the network. In the following some of the most important responsibilities of the MME are described.

- *Authentication and security:* Authentication is initiated by the MME at the time the UE registers in the network for the first time. The MME finds out the permanent identity of the UE either by asking from the former network that The UE was registered in or by asking from the UE itself. The MME also allocates a temporary identity to each UE which is called Globally Unique Temporary Identity (GUTI) for protecting the UE's privacy. Allocating the GUTI minimizes the need for sending the International Mobile Subscriber Identity IMSI. GUTI may be reallocated periodically in order to avoid unauthorized tracking of the UE.

- *Mobility management:* The MME follows the position of the UE and keeps track of UE's position. The MME defines the appropriate resource for the UE and asks the E-NodeB and the S-GW to allocate those resources to the UE. The MME has track of the UE at E-NodeB level if the UE is in connected mode and in Tracking Area (TA) level if the UE is in idle mode. The MME is also responsible for controlling of establishing and releasing resources to the UE based on UE activity mode. The other responsibility of MME is controlling the required signaling in handover procedure for a UE which is in active mode.

- *Managing subscription profile and service connectivity:* The MME has the responsibility for obtaining the user's subscription profile when the UE registers on the network. The MME will keep the subscription profile as long as the UE is served by the MME. The knowledge about this profile is very important for selecting the connection type between the PDN and the UE. Nevertheless, the MME can establish the basic establishment IP connectivity between the UE and the network including signaling between the UE and the E-NodeB and the S-GW in the control plane, and later, if there is a need for dedicated bearers, the MME can establish them.

Policy and Charging Resource Function (PCRF)

The PCRF is considered as a server that is located in the operator switching center, its responsibility being to control the policy and charging. In other words, its responsibility is service handling decision in terms of QoS and providing information for Policy Control Enforcement Function (PCEF). This information is called Policy and Charging Control (PCC) rules. PCC rules are sent by the PCRF whenever a new bearer needs to be set-up. As an example, establishing the default bearer at the time the UE connects to the network for the first time or when there is a need for dedicated bearer set-up can be mentioned.

Home Subscription Server (HSS)

The HSS is a database server which is located in the operator's premises. All the user subscription information is stored in the HSS. The HSS also contains the records of the user location. The HSS has the original copy of the user subscription profile. The subscription profile has information about the user such as the services that the user can use, the PDNs that the UE is allowed to connect to, and the networks that the UE can have the roaming to. The HSS is interacting with the MME, and it needs to be connected to all the MMEs in the network that controls the UE. When the UE is controlled by an MME, the HSS will keep the user location based on the information of that MME. The HSS also provides the information about the user profile to that MME. Upon receiving the user location from another MME, the HSS will cancel the former location information and update it based on the new MME information.

Service Domains

The service domain can be categorized to the IP Multimedia Sub-system (IMS) based operator services, non-IMS operator based services and the services which are not provided by the operator. In the first type of services, the services may be provided by using the Session Initiation Protocol (SIP). In the second type of services the operator may simply put a server in its network so that the UEs are able to connect with the server using an agreed protocol. These protocols should be supported by the applications the UEs are using.

2.4.2 The access network

In Figure 2, the whole architecture of E-UTRAN is depicted. As it is shown in the Figure 2, E-UTRAN is a mesh of eNodeBs that are connected to each other via X2 interface. The eNodeBs are also connected to the core network via S1 interface. The E-UTRAN architecture is considered to be flat because for normal user there is no centralized controller.

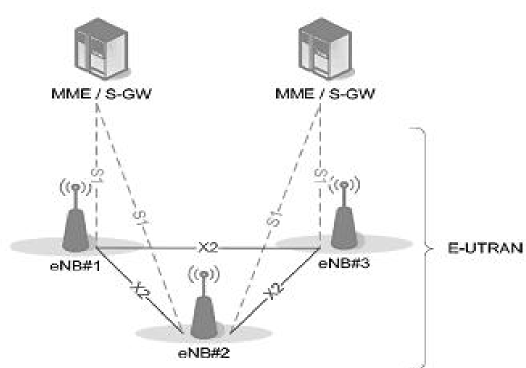

Figure 2: E-UTRAN protocol stack [5].

There are protocols that are running between eNodeBs and the UE. These protocols are called Access Stratum (AS) protocols. In this section we first explain the eNodeB and then summarize the responsibilities of the E-UTRAN with regard to radio-related functions.

The eNodeB

The eNodeB is a radio base station of a LTE network that controls all radio-related functions. These radio base stations are distributed in the network and each of them is placed near a radio antenna. Practically an eNodeB is a layer 2 bridge which is located between the UE and EPC. All the radio protocols that are used in the access link are terminated in the eNodeB. The eNodeB does ciphering/deciphering in the user plane as well as IP header compression/decompression. The eNodeB also has some responsibilities in the control plane such as radio resource management and performing control over the usage of radio resources. Controlling of radio resources can be categorized into resource allocation based on request, traffic prioritization and scheduling based on required QoS, and inspecting the situation of resource usage. The eNodeB also has an important role in Mobility Management (MM). It receives and analyzes the UE measurements regarding signal level, does some measurements itself and by comparing these two measurements decides to handover the UE. The eNodeB is required to exchange the signaling messages to the other eNodeBs as well as to the MME.

The ENodeB has a central role as regard to many other nodes. Several UEs can be served by the eNodeB when they are in eNodeB's coverage area, but each UE can be connected only to one eNodeB at a time. The eNodeB should have the connection to the other eNodeBs so that the UE has the possibility of handover. The eNodeB should also have the connection to MMEs and S-GWs. Some MMEs can be connected together and comprise a pool of MMEs. This can be also the case for S-GWs. The eNodeB can connect to these pools, although the UE gets its service only from one S-GW and is controlled only by one MME. In the following, responsibilities of the E-UTRAN regarding to radio related functions are described.

- *Radio Resource Management (RRM):* The RRM objective is to make the mobility feasible in cellular wireless networks so that the network with the help of the UE takes care of the mobility without user intervention. To enable mobility, the complexity both in the UE in terms of power consumption, processing power and cost, and in the network in terms of radio resource consumption and network topology increases. Therefore there is always a trade off between increasing the complexity and getting better performance. On the UE side, the main actions that are done to support seamless mobility are cell selection, measurement, handover, and cell reselection. In general, we can categorize all the responsibilities of RRM as dealing with the functions related to scheduling, admission control, radio bearer control and radio mobility control.

- *Header Compression:* IP packet headers can introduce huge overhead; therefore compressing the headers is a reasonable way for the efficient use of the radio interface. This responsibility is allocated to the PDCP which is one of the sublayers of the network layer. Header compression in PDCP is done by the use of the Robust Header Compression (ROHC) protocol. The importance of header compression in LTE is due to the use of packet switching and not circuit switching; therefore for the voice services to have the quality in packet switching domain close to the quality in the circuit-switched domain, header compression of IP/UDP/RTP headers is necessary. Otherwise enormous headers (compared to the small voice packets) will waste resources.

- *Security:* Providing security is necessary for avoiding unauthorized users to access the data. For the maintenance of security, two functions, ciphering and integrity protection, are used. Ciphering is applied both in control plane data like the Signaling Radio Bearer (SRB) and user plane data like the Data Radio Bearer (DRB) while integrity protection is done only in control plane data. Ciphering is a function that ensures that the third party avoids receiving the data streams. Integrity protection enables the receiver for detecting the packet insertion and replacement. The RRC always activate both of these functions together. In general, by security we mean that encryption is used for all the packet data sent over the radio interface.

- *Connectivity to the EPC:* The other function in the E-UTRAN is sending and receiving necessary signaling with the MME and establishing the bearers towards the S-GW.

  All of the above-mentioned functions are concentrated in the eNodeB as in LTE all the radio controller functions are gathered in the eNodeB. This concentration helps different protocol layers interact with each other better and will end up in decreased latency and increase in efficiency.

The User Equipment (UE)

The end user communicates using a UE. The UE can be a handheld device like a smart phone or it can be a device which is embedded in a laptop. The UE is divided into two parts: the Universal Subscriber Identity Module (USIM) and the rest of the UE, which is called Terminal Equipment (TE). The USIM is an application with the purpose of identification and authentication of the user for obtaining security keys. This application is placed into a removable smart card called a universal integrated circuit card (UICC). The UE in general is the end-user platform that by the use of signaling with the network, sets up, maintains, and removes the necessary communication links. The UE is also assisting in the handover procedure and sends reports about terminal location to the network.

2.5 Radio protocol architecture of E-UTRAN

The E-UTRAN protocols can be divided into user plane protocols and control plane protocols.

2.5.1 User plane

An IP packet is tunneled between the P-GW and the eNodeB to be transmitted towards the UE. Different tunneling protocols can be used. The tunneling protocol used by 3GPP is called the GPRS tunneling protocol (GTP). In Figure 3, the E-UTRAN protocol stack is depicted. The protocol stack consists of Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC) and Medium Access Control (MAC) sublayers that all terminate in the eNodeB in the network side. The PDCP is responsible for data protection during handover. RLC and MAC have different functionalities as discussed in the following sections. Unlike the PDCP, RLC and MAC both start a new session after handover with a new sequence numbers starting from zero.

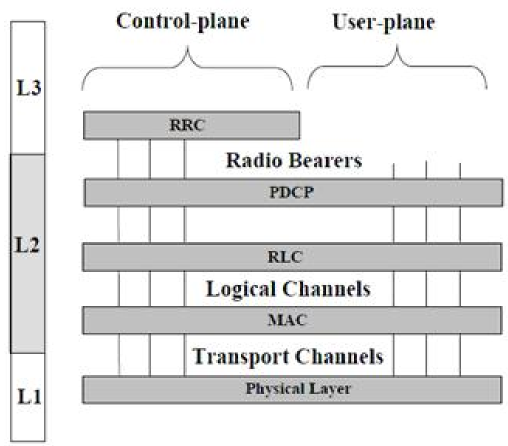

Figure 3: E-UTRAN [6].

Medium Access Control (MAC)

MAC is the lowest sublayer of Layer 2. It is located between the RLC layer and the physical layer. Logical channels connect MAC to the RLC and Transport channels connect MAC to the physical layer; therefore the main responsibility of the MAC layer is mapping the logical channels to the transport channels.

The MAC layer is located under the RLC layer in the protocol stack: therefore it receives packets from RLC in transmission and delivers packets to RLC in the reception. The main responsibility of the MAC layer can be concluded as multiplexing of RLC PDUs into Transport Blocks (TB) and then delivering them to the physical layer in transmission and demultiplexing TBs coming from physical layer to RLC PDUs and delivering them to the RLC layer in reception. Padding is also done by the MAC layer if the data does not completely occupy the MAC PDU. The other responsibilities of the MAC layer are as follows:

- MAC layer should measure and report the amount of traffic that passes this layer. The measurement and report is done for providing information about the amount of experienced traffic for the RRC layer.

- Uplink and downlink physical layer retransmission for error correction. The error correction is done through HARQ.

- *Scheduling functionality in the eNodeB:* there are available resources in a cell. The scheduling functionality in the eNodeB distributes these available resources between UEs in that cell. Also because a UE can have more than one radio bearer at a time, the scheduler should also distribute the resources in the cell between these bearers.

- *Scheduling information transfer:* the allocation of resources is only done when there is data to be sent or to be received. As mentioned earlier, scheduling is done in the eNodeB. In the downlink direction the eNodeB is aware of the amount of data in the buffer so it can allocate the radio resources easily, but in the uplink direction because the buffer is in the UE, sending a BSR from the UE to the eNodeB is necessary to inform the eNodeB about the amount of data in the buffer. The UE sends the Buffer Status Report (BSR) as regards the uplink buffer to the eNodeB. Then the eNodeB allocates uplink radio resources to the UE based on the BSR.

- *Random Access Procedure:* controlling the random access procedure is one of the important functionalities of the MAC layer. In two cases random access may be used:

  1. When the UE has some data to send but radio resources for uplink transmission are not allocated to the UE.
  2. When the UE is not time synchronized in the uplink.

- *Uplink Timing alignment:* Uplink timing alignment is very important in a sense that the uplink data of one UE does not overlap with the other UE's uplink data. Keeping the synchronization using uplink timing alignment when there is no data to be transferred, is a waste of resources. Therefore the UE is allowed to lose its synchronization even in RRC-CONNECTED mode. Then if some data appears, before transmission, the random access procedure will be used for the synchronization to be regained.

- *Discontinuous reception (DRX):* Generally the UE monitors downlink channels. By configuring the DRX functionality in RRC-CONNECTED mode, there will be a period of time that the UE is required to monitor the downlink channels and a period of time that the UE does not need to monitor. DRX functionality can prolong battery life. For setting the DRX parameters a trade off between battery saving and latency should always be considered.

Radio Link Control (RLC)

RLC is another sublayer of the data link layer. It is located between the PDCP and MAC. The communication between the RLC layer and the PDCP layer is done through the Service Access Point (SAP) and the communication of the RLC layer with the MAC layer is done through logical channels. Because RLC is located between the PDCP and MAC, it receives some PDCP PDUs from the PDCP layer in transmission time, reformats them and delivers them to the MAC layer. In reception time RLC receives RLC PDUs from MAC, reassembles them and sends them to the PDCP layer. The other functionality of RLC is reordering. In UMTS, reordering was a responsibility of the MAC layer, but in LTE this responsibility is appointed to RLC layer.

In RLC there are three transmission modes; Transparent Mode(TM), Unacknowledged Mode (UM), and Acknowledged Mode (AM). In the following these three modes are explained.

- *Transparent Mode (TM):* In TM no changes will be made to the PDUs that pass this entity and no overhead is added to these PDUs. Not adding any overhead means RLC SDUs are mapped directly to RLC PDUs and vice versa. The messages like RLC messages that do not need any RLC configuration can use TM modes; otherwise TM Mode is not used for user plane data at all. TM RLC mode provides unidirectional data transfer either for transmission or for reception.

- *Unacknowledged Mode (UM):* unacknowledged mode as its name indicates does not have any retransmission. Therefore using the UM entity provides less delay and more error probability. Real time applications such as VoIP that are sensitive to delay but can tolerate the errors use UM mode. similar to TM RLC the data transfer in UM RLC is unidirectional. In the following the main functions of UM RLC are discussed.

  1. **Segmentation and concatenation of RLC SDUs:** segmentation and concatenation are responsibilities of the RLC layer. These two functions are performed on RLC SDUs that are coming from an upper layer to make RLC PDUs. The size of RLC PDUs can be different depending on the radio channel conditions and available resources that are notified by MAC.
  2. **Reordering, duplicate detection, reassembly:** upon receiving RLC PDUs, UM RLC entity reorders them if they are not received in the correct sequence. The RLC PDU which is out of sequence is kept in the buffer until all the former RLC PDUs are received and are delivered to the higher layer. Also by checking the sequence number all the duplicated RLC SDUs can be identified and deleted. Receiving duplicated PDUs is usually due to misinterpreting of Hybrid Automatic Repeat reQuest(HARQ) ACKs as HARQ NACKs in the MAC layer. Deleting

duplication is part of reordering process. Reassembly is only done when all the segments of RLC SDUs are available and it is done using the stored RLC PDUs.

- *Acknowledged Mode (AM):* AM RLC is the only mode that provides bidirectional data transfer. The prominent difference of AM RLC with UM RLC is retransmission; therefore all the functions performed by UM RLC are applicable for AM RLC as well. For instance, concatenation and segmentation of RLC SDUs is done on the transmitter side to form RLC PDUs and reassembling of RLC SDUs from the received RLC PDUs is done on receiver side. AM RLC mode has retransmission using an Automatic Repeat reQuest (ARQ) to support error-free transmission. Since retransmission can correct the errors in transmission, AM RLC is a suitable mode for non-realtime applications that are sensitive to errors and can tolerate delays. Besides the functionalities that are common between AM RLC and UM RLC, AM RLC has other functions which are explained in the following.

  1. **Retransmission and resegmentation:** Retransmission should be done only for the RLC SDUs that have not been received at the receiver side. Therefore the receiver sends a status report comprising the ACK and NACK information of RLC PDUs. Then the transmitter based on this information decides which RLC PDUs should be transmitted again. In general, a transmitter is able to send two types of RLC PDUs: RLC data PDUs that are received from the upper layer and RLC control PDUs that are generated in RLC entity. There is a flag in the AM RLC header to distinguish RLC data PDUs and RLC control PDUs.

  2. **Polling, status report and status prohibit:** the transmitter side of AM RLC can request for a status report from the receiver side. This function is called polling and can be done by setting an allocated flag in the AM RLC header to one. By setting this flag the transmitter side can continuously receive the status report. The transmitter side can use the status report to identify the RLC data PDUs that need to be retransmitted and retransmit them. When the receiver side receives the polling then it checks its buffer and sends the status to the transmitter at the first possible transmission opportunity.

     In status report transmission, a trade off between delay and radio resources should be considered. In order to decrease the delay, frequent transmission of the status report is desired, but on the other hand, frequent transmission of status report wastes valuable radio resources. Also more radio resources might be wasted, if before reception of a retransmission that has started due to a former status report a new retransmission due to a new status report initiates. This is considered as unnecessary retransmission and wastes radio resources. To avoid the unnecessary retransmission mentioned above and to provide good control over sending the status reports, in AM RLC there is a “status prohibit” function which does not let a new status report initiate before the retransmission based on the former retransmission has been completed.

Packet Data Convergence Protocol (PDCP)

PDCP stands for Packet Data Convergence Protocol. It is one of the sub layers in the Data Link layer. The PDCP protocol terminates in the eNB from one side and in the UE from the other side, and it also acts both in the user plane and control plane. The PDCP receives Radio Resource Control messages in the control plane and processes IP packets in the user plane. Depending on the radio bearer, the main functions of the PDCP layer are header compression, security (integrity protection and ciphering), and support for reordering and retransmission during handover. There is one PDCP entity per radio bearer.

The PDCP includes some entities. Each RB (DRB and SRB except SRB0) is associated with one PDCP entity. Then each PDCP entity is associated with one or two (one for each direction) RLC entities depending on RB characteristics (uni-directional or bi-directional) and RLC mode. Several PDCP entities may be defined for a UE. Each PDCP entity carries the data of one radio bearer. A PDCP entity is associated either to the control plane or the user plane depending on which radio bearer it is carrying the data for. The PDCP provides services both to the upper layer and to the lower layer. These services are as follows.

PDCP Services

1. Transfer of user plane and control plane data
2. Header compression
3. Ciphering
4. Integrity protection.

There are also some services that the PDCP provides to the lower layer. These services are:

1. Acknowledged data service, including indication of successful delivery of PDCP PDUs
2. Unacknowledged data transfer service
3. In-sequence delivery, except at re-establishment of lower layers
4. Duplicate discarding, except at re-establishment of lower layers.

PDCP Functions

Besides the services there are also some functions which are used by the PDCP. These functions are:

1. Header compression and decompression of IP data flows using the ROHC protocol
2. Transfer of data (user plane and control plane)
3. Maintenance of PDCP SNs
4. In-sequence delivery of upper layer PDUs at re-establishment of lower layer
5. Duplicate elimination of lower layer SDUs at re-establishment of lower layers for radio bearers mapped on RLC AM
6. Ciphering and deciphering of user plane data and control plane data
7. Integrity protection and integrity verification of control plane data
8. Timer based discard
9. Duplicate discarding.

PDCP PDU Types

In LTE there are two different types of PDCP PDU: PDCP Control PDUs and PDCP Data PDUs. PDCP Control PDUs are used only in user plane. They carry either ROHC feedback or PDCP status reports in the case of handover. PDCP Data PDUs are used for both control plane and user plane data.

PDCP PDU Formats

A PDCP PDU is a bit string, that its length being a multiple of 8 bits, which is called byte aligned. In Figure 4, the bit strings are represented by tables in those each line comprises of 8 bits. Basically the most significant bit is the left most bit in the first line and the least significant bit is the right most bit in the last line.

The PDCP SDUs are also bit strings that are byte aligned (they have the length of a multiple of 8 bits). A PDCP PDU includes at least one PDCP SDU that starts from the first bit.

- *Control plane PDCP Data PDU:* Figure 4 shows the format of PDCP Data PDUs for SRBs in control plane. In the example, PDCP SN is 5 bits long.

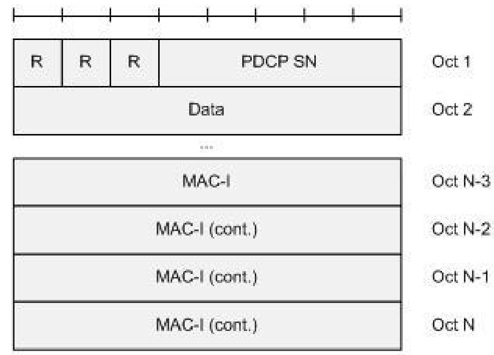

Figure 4: PDCP Data PDU format for SRBs [7].

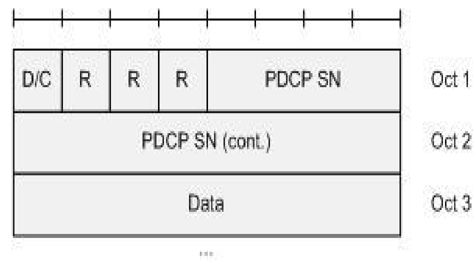

Figure 5: PDCP Data PDU format for DRBs using a 12-bit SN [7].

- *User plane PDCP Data PDU with long PDCP SN (12 bits):* In Figure 5, the format of PDCP Data PDU with the 12-bit sequence number length is depicted. The format is used for DRBs mapped on AM RLC or UM RLC. The same format with the 7-bit sequence number is used for DRBs mapped on RLC UM (Figure 6).

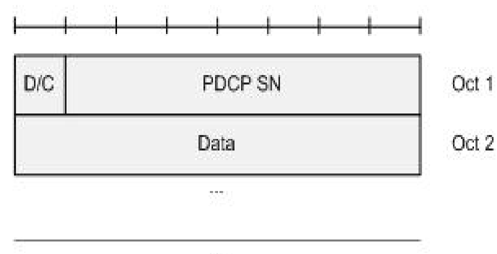

Figure 6: PDCP Data PDU format for DRBs using 7-bit SN [7].

- *PDCP Control PDU for interspersed ROHC feedback packet:* In Figure 7, the format of the PDCP Control PDU which carries an ROHC feedback packet is depicted. The format is used for DRBs mapped on both AM RLC and UM RLC.

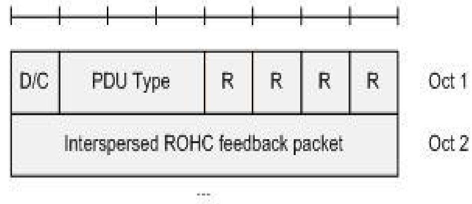

Figure 7: PDCP Control PDU format for interspersed ROHC feedback packet [7].

- *PDCP Control PDU for PDCP status report:* In Figure 8, the format of PDCP Control PDU that carries one PDCP status report is depicted. The format is applicable for DRBs mapped on RLC AM.

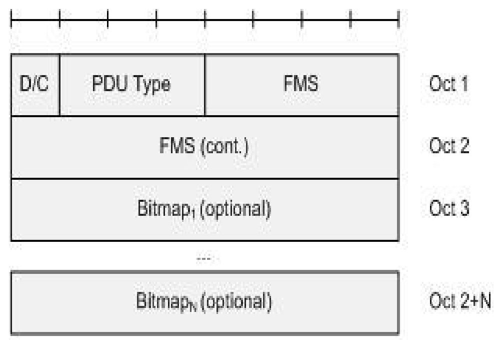

Figure 8: PDCP Control PDU format for PDCP status report [7].

The different fields in Figures 4 to 8, are: R field which is a sign for reserved bit, Message Authentication Code for Integration (MAC-I) which carries a message authentication code, Count which is maintained for ciphering and integrity purposes, D/C that shows if the PDU is a Control PDU or a Data PDU, First Message PDCP SN (FMS) that contains the PDCP SN of first missing PDCP SDU, Bitmap with the information about the PDCP SDUs in need for retransmission, PDCP SN that shows the sequence number for PDCP, and the Data field which has a variable length and can contain uncompressed or compressed PDCP SDUs.

2.5.2 Control plane

In Figure 3 the protocol stack in the control plane is shown. The common protocols between the control plane and the user plane have the same functionality except that for the control plane protocols there is no header compression. In the access stratum protocol stack and above the PDCP, there is the RRC protocol which is considered as a “Layer 3” protocol. RRC sends signaling messages for establishing and configuring the radio bearers of all lower layers in access stratum. A UE has two different Radio Resource Control (RRC) states that are RRC-IDLE and RRC-CONNECTED. When a UE is in RRC-IDLE mode it decides about the cell that it is camping on. The first decision is called cell selection and all the following decisions are called cell reselection. From the paging channel the UE in RRC-IDLE mode can receive the notification of incoming calls. System information parameters are necessary for cell reselection. In the rest of this section RRC and its messages are discussed.

Radio Resource Control (RRC)

Among the control messages that are exchanged between the E-UTRAN and the UE, Radio Resource Control messages have the highest share. Compared to the UTRAN, RRC in the E-UTRAN has fewer messages and less redundancies in the messages, therefore it is simpler. Also UE states in E-UTRAN compared to the UTRAN are simpler. The UE has only two states, RRC-CONNECTED and RRC-IDLE. In the rest of this section the main responsibilities of the RRC protocol are explained.

– *System information broadcast.* System information has the information of both non-Access Stratum (NAS) and AS. System information elements will be grouped together and create Master Information Block (MIB) and System Information Block (SIB). The information in the MIB is very important. Therefore MIB is sent through BCH. SIBs are numbered from SIB1 to SIB11. Each of them contains specific information. For instance SIB1 has the information related to access such as cell selection and scheduling information. As the other example, SIB10 has the information about the Earthquake and Tsunami Warning System (ETWS) primary alarm and SIB11 has information for the ETWS secondary alarm.

– *Paging.* The prominent usage of paging is to page the UEs that are in RRC-IDLE. Paging can also be used to notify UEs both in RRC-IDLE and RRC-CONNECTED modes about system information changes or SIB10 and SIB11 transfers.

– *Controlling cell selection and cell reselection.* UE selects a cell and camps on it based on the measurements it has done in IDLE mode and information of the SIBs.

– *Handling RRC connection between E-UTRAN and UE.* There are three RRC messages that are handling the connection between the E-UTRAN and the UE. The connection set up can start by a request from the UE in NAS. There are five different reasons defined for the request that are emergency, high priority access, mobility terminated access, mobile originating signaling and mobile originating data. After the UE sends an RRC connection request message, it receives the RRC connection setup message from the eNodeB. Then the UE sends an RRC connection setup complete message. If the RRC connection setup procedure is successful, the UE moves from IDLE mode to the CONNECTED mode.

– *Security.* There are some security keys that are used in Access Stratum (AS), to provide security both for user data and RRC control signaling. After completion of the security procedure, the EPC and UE share the same master key. The eNodeB does not have the master key, but the necessary keys for the UE and eNodeB communications that are derived from master key are delivered to the eNodeB. The eNodeB keeps the keys as long as the UE is connected, but will delete them when the UE moves to IDLE mode or performs a handover to another eNodeB. Upon the handover and reestablishment, the keys will be changed. Ciphering and integrity algorithms are also changed upon handover.

– *Handling point to point radio bearers.* The RRC connection reconfiguration procedure is done using two messages: RRC connection reconfiguration and RRC connection reconfiguration complete. The main goal of this procedure is keeping and releasing radio bearers. RRC connection reconfiguration can only release the data radio bearers and signaling radio bearers cannot be released by this procedure.

– *Controlling UE measurement reporting.* In the RRC reconfiguration message there is a parameter for configuring a measurement report. The first configuration measurement parameter is a measurement object based on which the UE should perform a handover. As an example intra-frequency measurement and inter-frequency measurement can be two different objects. Intra-frequency measurement means performing a measurement on the same carrier frequencies as the frequencies in the downlink of the serving cell while inter-frequency measurement means performing measurement in the frequencies different from the downlink frequencies of the serving cell.

  The second parameter is reporting configuration which configures the reporting criteria and the reporting format. Reporting criteria shows how the measurement report should be triggered by the UE. There are two types of triggering; either periodically or based on an event. Reporting format contains some numbers. For example, one of these numbers shows the number of cells that the UE should do measurement towards them and include them in its measurement report. Regarding criteria, there are different events which are addressed here. Note that B1 and B2 are the events for inter-RAT measurement.

  Event A1: serving cell becomes better comparing to the absolute threshold

  Event A2: serving cell becomes worse comparing to the absolute threshold

  Event A3: neighboring cell becomes better than the serving cell to the amount of an offset

  Event A4: neighboring cell becomes better than the absolute threshold

Event A5: serving cell becomes worse than the absolute threshold 1 and neighboring cell becomes better than absolute threshold 2

Event B1: neighboring cell becomes better that the absolute threshold

Event B2: serving cell becomes worse than the absolute threshold 1 and neighboring cell becomes better than the absolute threshold 2.

– *Handover.* The handover is triggered by the eNodeB, based on the received measurement reports from the UE. Handover is classified in different types based on the origination and destination of the handover. The handover can start and end in the E-UTRAN, it can start in the E-UTRAN and end in another Radio Access Technology (RAT), or it can start from another RAT and end in E-UTRAN. Handover types are categorized.

  – intra-frequency intra-LTE handover
  – inter-frequency intra-LTE handover
  – inter-RAT towards LTE handover
  – inter-RAT towards UTRAN handover
  – inter-RAT towards GERAN handover
  – inter-RAT towards cdma2000 system handover.

  From the handover classes mentioned above we shortly explain intra-LTE handover.

Intra-LTE handover

In intra-LTE handover, both the origination and destination eNBs are within the LTE system. In this type of handover, the RRC connection reconfiguration message acts as a handover command. The interface between eNodeBs is an X2 interface. Upon handover, the source eNodeB sends an X2 handover request message to the target eNodeB in order to make it ready for the coming handover.

3 Handover

3.1 Introduction

Mobility is an important feature in wireless networks. Increasing the speed of vehicles on one side and the need to use Internet almost everywhere at any time emphasizes on the necessity and the importance of mobility in wireless networks. Mobility at high speed is a challenge, and LTE as long term evolution has promised more than former technologies to overcome this challenge. To accomplish this purpose, minimum possible delay and packet loss in voice transmission and reliability in data transmission are desired. Therefore optimizing the handover procedure to get the required performance is considered as one important issue in mobile networks.

As the UE moves it can face different propagation conditions and different interference levels. It might happen that the cell which is serving the UE is not the best cell anymore and the UE needs to be handed over to another cell. For this to happen, while the UE is connecting to the serving cell, the UE needs to continuously monitor neighboring cells. The cell that serves the UE is called the source cell or the serving cell and the cell that the UE is handed over to is called the target cell. In LTE, we can distinguish between two different mobility modes: mobility in IDLE mode and mobility in CONNECTED mode. When the UE is in IDLE mode and changes the cell, the process is called cell reselection, and while the UE is in CONNECTED mode and changes cell, the process is called handover. The network controls the UE transitions from IDLE mode to CONNECTED mode and vice versa.

The handover in LTE is hard handover, instead of soft handover in WCDMA networks, because the UE connection to the old (source) eNB breaks before the new connection to the new (target) eNB is established; there will be therefore a handover interruption time in the user plane. As data loss should be avoided in handovers, data forwarding is developed for LTE networks. As mentioned earlier, there is no soft handover in LTE, as well as any centralized controller node. Therefore the responsibility of handling data in handover is assigned to the eNB. In more detail, data buffering and protection during handover due to user mobility in the E-UTRAN is assigned to the PDCP layer in the eNB, which is a sub layer of the data link layer. Handover is one of the procedures followed by the UE to provide seamless mobility. The handover is beneficial for the end-user although it increases the complexity of the network and the UE. Radio interface resource consumption and network topology can be considered as network complexity. Without handover the end-user cannot experience seamless mobility and may lose connectivity while leaving a cell and entering another cell. The objective in networks like LTE is providing seamless mobility for the user, and at the same time keeping the network management simple.

In this chapter after looking closely into the handover procedure, the necessary definitions related to handover are brought out to provide an overview about handover.

After this data forwarding as one of the steps in the handover procedure with the possibility of handover performance improvement is discussed and finally the former works in this area are investigated to provide a good background about the topic as well as revealing the promising areas for performing research in the future.

3.2 Handover procedure

EPC is not involved in Handover (HO) procedure and all the necessary messages are directly exchanged between the eNBs. The handover procedure is illustrated in Figure 9 and is considered in the following.

1. Based on the area restriction information, the source eNB configures the UE measurement procedure.

2. MEASUREMENT REPORT is sent by the UE after it is triggered based on some rules.

3. The decision for handover is taken by the source eNB based on MEASUREMENT REPORT and RRM information.

4. HANDOVER REQUEST message is sent to the target eNB by the source eNB containing all the necessary information to prepare the HO at the target side.

5. The target eNB may perform an Admission Control dependent on the received E-RAB QoS information. Performing admission control is to increase the likelihood of a successful HO, in that the target eNB decides if the resources can be granted or not. In case the resources can be granted, the target eNB configures the required resources according to the received E-RAB QoS information then reserves a Cell Radio Network Temporary Identifier (C-RNTI) and a RACH preamble for the UE.

6. The target eNB prepares HO and then sends the HANDOVER REQUEST ACKNOWLEDGE to the source eNB. There is a transparent container in the HANDOVER REQUEST ACKNOWLEDGE message which is aimed to be sent to the UE as an RRC message for performing the handover. The container includes a new C-RNTI, target eNB security algorithm identifiers for the selected security algorithms, may include a dedicated RACH preamble, and possibly some other parameters like RNL/TNL information for the forwarding tunnels. If there is a need for data forwarding, the source eNB can start forwarding the data to the target eNB as soon as it sends the handover command towards the UE.

   Steps 7 to 16 are designed to avoid data loss during HO:

7. To perform the handover the target eNB generates the RRC message, i.e RRCConnectionReconfiguration message including the mobilityControlInformation. This message is sent towards the UE by the source eNB.

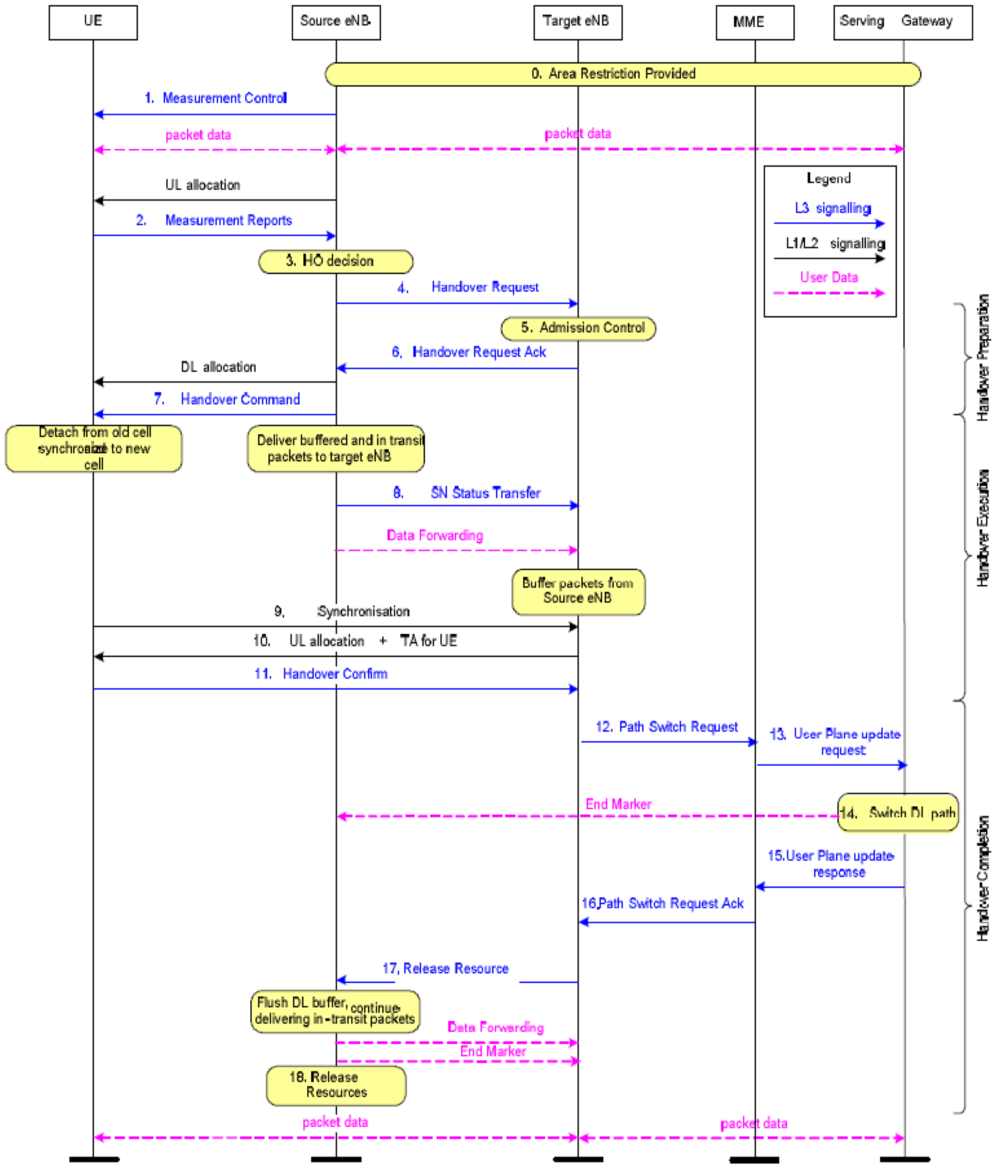

Figure 9: Handover Procedure [8].

8. The SN STATUS TRANSFER message is sent by the source eNB to the target eNB. In that message, the information about uplink PDCP SN receiver status and the downlink PDCP SN transmitter status of E-RABs are provided. The PDCP SN of the first missing UL SDU is included in the uplink PDCP SN receiver status. The next PDCP SN that the target eNB shall assign to the new SDUs is indicated by the downlink PDCP SN transmitter status.

9. After reception of the RRCConnectionReconfiguration message including the mobilityControlInformation by the UE, the UE tries to perform synchronization to the target eNB and to access the target cell via RACH. If a dedicated RACH preamble was assigned for the UE, it can use a contention-free pro-

cedure, otherwise it shall use a contention-based procedure. In the sense of security, the target eNB specific keys are derived by the UE and the selected security algorithms are configured to be used in the target cell.

10. The target eNB responds based on timing advance and uplink allocation.

11. After the UE is successfully accessed to the target cell, it sends the RRC-ConnectionReconfigurationComplete message for handover confirmation, The C-RNTI sent in the RRCConnectionReconfigurationComplete message is verified by the target eNB and afterwards the target eNB can now begin sending data to the UE.

12. A PATH SWITCH message is sent to MME by the target eNB to inform that the UE has changed cell.

13. UPDATE USER PLANE REQUEST message is sent by the MME to the Serving Gateway.

14. The Serving Gateway switches the downlink data path to the target eNB and sends one or more "end marker" packets on the old path to the source eNB to indicate no more packets will be transmitted on this path. Then U-plane/TNL resources towards the source eNB can be released.

15. An UPDATE USER PLANE RESPONSE message is sent to the MME by the Serving Gateway.

16. The MME sends the PATH SWITCH ACKNOWLEDGE message to confirm the PATH SWITCH message.

17. The target eNB sends UE CONTEXT RELEASE to the source eNB to inform the success of handover to it. The target eNB sends this message to the source eNB after the PATH SWITCH ACKNOWLEDGE is received by the target eNB from the MME.

18. After the source eNB receives the UE CONTEXT RELEASE message, it can release the radio and C-plane related resources. If there is ongoing data forwarding it can continue [8].

3.3 Handover-related definitions

In the former section, the handover procedure was covered. In this section, we describe some of the definitions such as seamless handover, lossless handover, control plane and user plane handover handling and handover delay.

3.3.1 Seamless Handover

Seamless handover is the main concern in all systems having mobility. The objective of seamless handover is to provide a given QoS during the process of migration from one domain to another and also to hide any difference between the normal services offered within a domain and during a migration from the user. In LTE Seamless handover is applicable for all the radio bearers carrying control plane data, and user plane data mapped on RLC-UM. These two types of data are tolerant for losses but less tolerant for delay, therefore seamless handover should minimize the complexity and delay although some SDUs might be lost.

In the seamless handover, PDCP entities are reset, and because there is no security reason to keep COUNTS values, due to using of new keys, after handover, COUNT values are set to zero. On the UE side, all the PDCP SDUs that have not been transmitted yet will be sent to the target cell after handover. In the eNodeB, PDCP SDUs for those transmissions have not been started can be forwarded via X2 towards the target eNB. Unacknowledged PDCP SDUs will be lost. Therefore handover complexity is minimized.

3.3.2 Lossless Handover

Lossless handover means that no data should be lost during handover. This is achieved by performing retransmission of PDCP PDUs for which reception has not been acknowledged by the UE before the UE detaches from the source cell to make a handover. In lossless handover, in-sequence delivery during handover can be ensured by using PDCP Data PDUs sequence numbers. Lossless handover can be very suitable for delay-tolerant services like file downloads that the loss of PDCP SDUs can enormously decrease the data rate because of TCP reaction. Lossless handover is applied for radio bearers that are mapped on RLC-AM. In lossless handover, on the UE side the header compression is reset because its context is not forwarded from the source eNB to the target eNB, but the PDCP SDUs' sequence numbers and the COUNT values are not reset. In the following we consider an example to clarify lossless handovers both in uplink and downlink transmissions.

For having a lossless handover in uplink the UE will retransmit all the PDCP SDUs that are in the PDCP retransmission buffer. In the example demonstrated in Figure 10, the PDCP PDUs with a sequence number from 1 to 5 are transmitted to the eNB but the PDCP PDUs with sequence numbers 3 and 5 do not reach the destination. After the handover, the UE transmits all the packets that have not been acknowledged before handover. Since in the example above the only packets that are acknowledged are packets 1 and 2, packets 3, 4, and 5 will be retransmitted. For having in-sequence delivery in uplink, the source eNB will send the packets that have come in a correct order to the gateway and forwards the ones that are not in-order to the target eNB after decompression. Then the target eNB can reorder these packets and the retransmitted packets based on the PDCP SN and afterwards

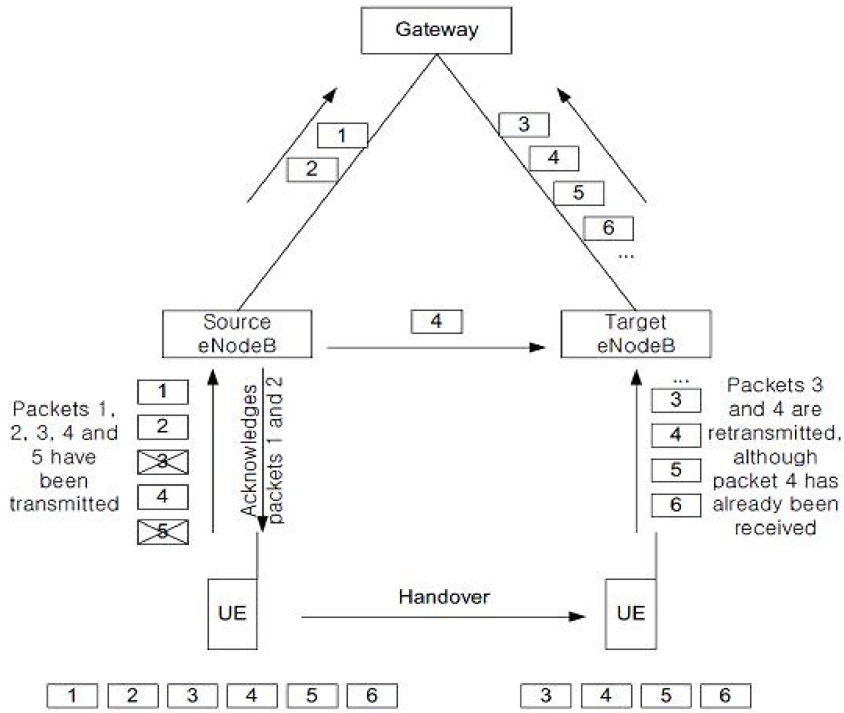

Figure 10: Lossless handover in uplink [5].

can pass them to the GW. Figure 11, shows lossless handover in downlink.

3.3.3 The UE in handover

The UE is expecting to receive downlink packets in an ascending order and if the UE finds a gap in packet reception, it will conclude that some packets have been lost. In this case, the UE delivers the other in-sequence packets to the higher layer and does not delay sending already received packets for potential retransmission. As was mentioned earlier, the handover changes the serving cell of a UE in RRC-CONNECTED, and it is different from cell reselection in that the UE is in RRC-IDLE mode. It was also mentioned that handover in LTE is NW controlled handover which is assisted by the UE and the E-UTRAN commands the UE to start, modify or stop the measurements. The assistance of the UE in the handover procedure is in measurement. The measurement can be distinguished in two different categories when making handover in LTE networks:

- Intra-frequency E-UTRAN measurements
- Inter-frequency E-UTRAN measurements.

After the measurement is done, it should be reported to the eNB. For the reporting three different criteria can be used:

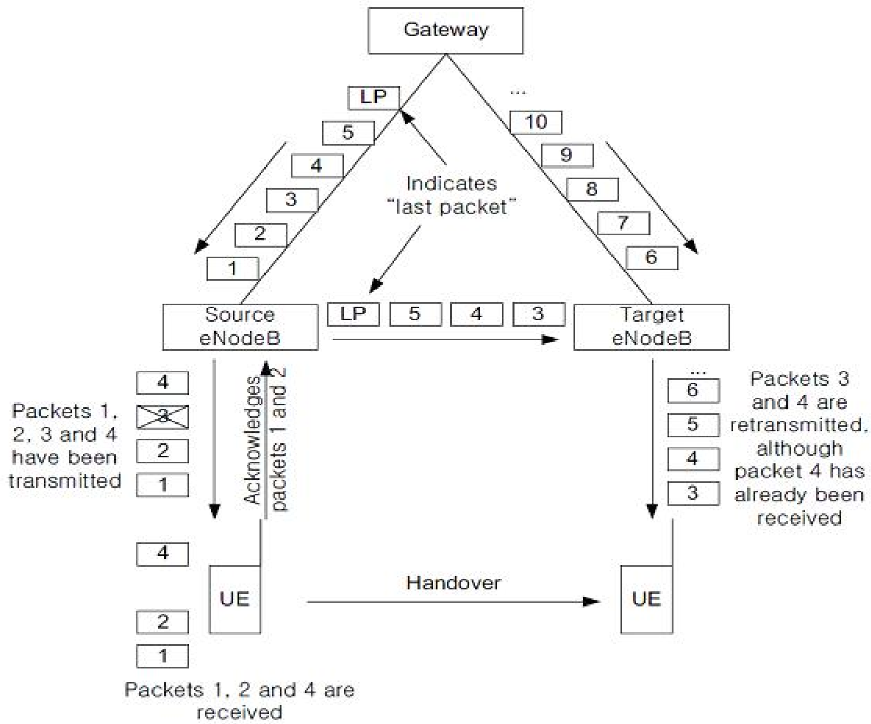

Figure 11: Lossless handover in downlink [5].

– event triggered reporting

– periodic reporting

– event triggered periodic reporting.

The UE uses RACH to access the target cell. If within a certain time the RACH procedure is not successful, radio link failure recovery using the best cell will be initiated by the UE.

3.3.4 Control plane handling in handover

The EPC is not involved in handover procedure and all the messages are transferred between eNBs. Even releasing the resources in the source side during the handover completion time is triggered by the eNB. When relays are deployed in the network, their DeNB are responsible for sending the relay information to the MME in case of S1-based handover and to the target eNB in case of X2-based handovers. Using the X2 and S1 proxy functionality, the DeNB is explicitly aware of the UEs that are connected to the RN.

3.3.5 User plane handling in handover

In the HO preparation phase, U-plane tunnels are established between the source and the target eNBs. For each E-UTRAN Radio Access Bearer (E-RAB), one tunnel is established for uplink data forwarding and one tunnel for downlink data forwarding. If there are relays deployed, the tunnels between the RN and the target eNB are established via the DeNB. In the handover execution phase, the user data is forwarded from the source eNB to the target eNB. Forwarding continues until all the packets that have not been acknowledged by the UE, transfer to the target eNB to be retransmitted to the UE. In the HO completion phase, the PATH SWITCH message is sent to the MME to inform that the UE has accessed the target eNB, then the MME sends a USER PLANE UPDATE REQUEST message to the Serving Gateway and the serving Gateway changes the U-plane path from the source eNB to the target eNB. Before the path switch still some packets will be sent to source eNB that should be forwarded to the target eNB.

- For RLC-AM bearers, in-sequence delivery and avoiding duplication can be obtained by allocating continuous PDCP SNs. The source eNB will send the last allocated PDCP SN to the target eNB and the target eNB continues allocating PDCP SN to the packets that have come whether from the source eNB or the serving gateway and do not have a PDCP SN. To detect duplication, a window-based mechanism is needed both in the UE and the target eNB. The target eNB re-transmits and prioritizes all downlink PDCP SDUs forwarded by the source eNB that means the data with PDCP SNs from X2 should be sent by the target eNB before data from S1. PDCP SDUs of which the reception was acknowledged through PDCP SN based reporting by the UE are not retransmitted. All uplink PDCP SDUs starting from the first PDCP SDU following the last consecutively confirmed PDCP SDU, i.e. the oldest PDCP SDU that has not been acknowledged at RLC in the source, are retransmitted in the target eNB by the UE, except for the PDCP SDUs of which the reception was acknowledged through PDCP SN based reporting by the target. The security synchronization is obtained by using Hyper Frame Number (HFN), which is provided by the source eNB to the target eNB.

- For RLC-UM bearers unlike for RLC-AM, the PDCP SN and HFN are not continued, instead they are reset in the target eNB and there is no PDCP SDU retransmission in the target eNB. Forwarded PDCP SDUs from the source eNB are prioritized in the target eNB, meaning that the data received with PDCP SNs from X2 should be sent before sending data received from S1. Upon changing the path (path switch), one or more end marker packets should be sent towards the old path to indicate that no new data packets will be transmitted through the old path. These end markers can also show the end of data forwarded packets in the target eNB, as they will be sent after all the data packets through the forwarding tunnel. The target eNB receiving these end markers know that all the packets from the forwarded path are received and after sending them to the UE, can start sending the other packets that have come through a new S1 path.

3.3.6 Handover delay

When the UE drives around the corner in an urban area there might be a high probability that the signal level from the source cell degrades fast while the signal level from the target cell improves slowly. In this case, for maintaining the service, a fast handover is inevitable. The UE should receive the handover command before the signal from the source eNB degrades too much. Therefore minimizing the network delay in response to the measurement report done by the UE is crucially necessary for providing a reliable handover. In Figure 12 all necessary timings for handover are shown.

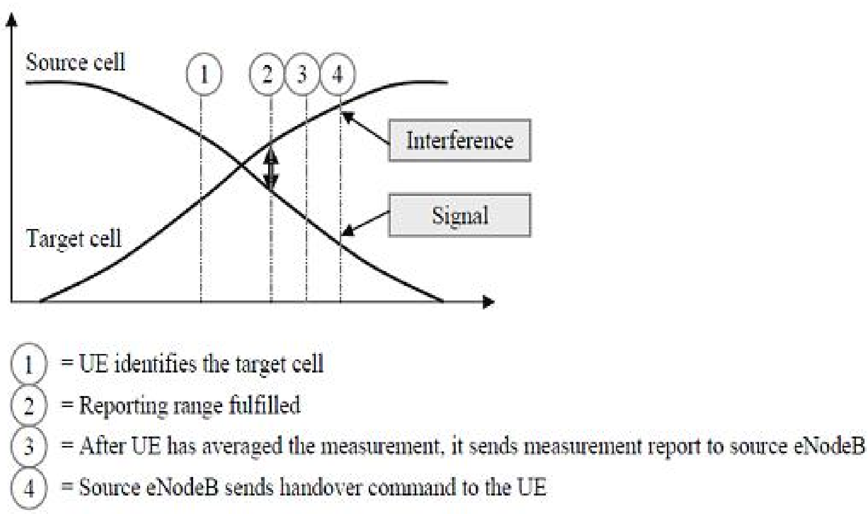

Figure 12: Handover Timing [6].

3.4 Data forwarding

Data forwarding is one of the steps in the handover procedure that can be configured in a network to ensure a lossless handover. If data forwarding is configured, during handover, the source eNB forwards all the downlink PDCP SDUS, which have not been acknowledged by the UE to the target eNB. Before the path switch, some new packets may also arrive in the old S1 path that should also be forwarded to the target eNB. The target eNB will then start transmitting these packets to the UE without waiting to receive all the packets. The forwarding is done at the PDCP level. At the RLC level, all the remaining downlink RLC PDUs will be discarded without being forwarded to the target eNB.

Although forwarding is possible both on RLC-AM and RLC-UM bearers, only the PDCP RLC-AM bearers can support SN continuity on the PDCP level. For the PDCP RLC-UM bearers, the sequence numbers are reset, and the data on the lower layer is discarded at the re-establishment procedure. The "in-order delivery and duplicate elimination" function at the UE PDCP layer promises in-sequence delivery of downlink PDCP SDUs in the downlink and the "in-order delivery and duplicate elimination" function at the target eNB PDCP layer promises in-sequence delivery of uplink PDCP SDUs in the uplink.

The PDCP Status Report has an important role in data forwarding, having the information about the PDCP SDUs that have already been correctly received to avoid retransmission. It also has the request of retransmission for the PDCP SDUs that have not been received correctly, besides the PDCP SDUs that have been received correctly with a failed header decompression. The PDCP status report contains bitmap and First Missing Sequence (FMS). Bitmap shows which PDCP SDUs need to be retransmitted and FMS is used as a reference that shows the sequence number of the first missing PDCP SDU. The use of bitmap and FMS assures in-sequence delivery of PDCP SDUs.

3.5 The main criteria for designing handovers

The handover in LTE is initiated by the network and assisted by the UE. For the handover, RSRP and RSRQ measurements are made in the UE and they are sent to the eNB regularly. There are also some predefined handover conditions or threshold definitions in the network for triggering the handover procedure as well as some goals regarding handover design and optimization such as decreasing the total number of handovers in the whole system by predicting the handover, decreasing the number of ping pong handovers, and having fast and seamless handover. To reach these goals and to response to the drivers like competition and end-user demands, further improvements in the terms of latency, user data rate, system capacity and cost should be considered. In the following the main criteria for designing handovers are discussed.

- *Minimizing the number of handover failures:* Minimizing the number of handover failures is necessary for avoiding call termination during handover time and letting the UE be connected and continue conversation or down/upload data during and after handover.

- *Minimizing the number of unnecessary handovers:* Minimizing the number of unnecessary handovers is to guarantee the communication quality and to avoid suffering from the large number of the handovers that increases the signaling load in the network. Minimizing the number of handovers also avoids increasing the risk of call drops related to interruption in handover.

- *Minimizing handover delay:* Due to the fact that handover in LTE is a hard handover and interruption might be noticed by the user; a fast handover is needed not to let the user experience service degradation or interruption.

– *Maximizing the amount of time that the UE is connected to the best cell:* Handover is performed to have the UE connected to the best cell. Achieving this goal will be easier if the handover is designed in a way that prolongs the amount of time that the UE is connected to the best cell.

– *Minimizing the impact of handover on system and service performance:* Minimizing the impact of handover on system and service performance can be obtained by optimizing the handover procedure. Some of the goals mentioned above are in contradiction to each other. For example, maximizing the amount of time that the UE is connected to the best cell increases the number of handovers. Therefore making a tradeoff between these goals to approach the required ultimate goal will be necessary. As is brought out in the following, the handover has different parameters. Setting these parameters to the optimal values provides handovers with desirable performance. The decision to trigger a handover is also very important. In LTE the triggering is usually based on RSRP and RSRQ measurement and some other parameters to improve the performance. In the following the most important ones are addressed.

– *Handover initiation threshold level:* If handover initiation threshold level decreases, the handover will be triggered faster and if it increases the probability of having late handover will increase.

  – *Hysteresis margin* is for avoiding ping pong handovers, and it is defined as a margin that link quality of target cell should be greater than the link quality of source cell plus this margin.
  – *Time to Trigger(TTT)* is the amount of time that triggering requirements should be fulfilled during this time, then handover can be initiated.
  – *Shape and the length of averaging window* should be selected carefully in the handover decision for minimizing the effect of the channel variation due to fading.

3.6 Former research

After having an overview of the handover, the following section covers the former studies regarding to improving handover performance and data forwarding.

In [9], a simulation study for some handover algorithms for enhancing the performance of the LTE systems is presented. It is concluded that the handover in the LTE system is complicated and imperfect, and doing research to optimize handover algorithms in the LTE system is a necessity. In LTE and LTE-A systems the handover is hard handover with its shortcomings like high handover outage probability and large delay. Using semi soft-handover instead of hard handover is proposed for enhancing handover performance.

In contrast to [9], an overview of the LTE intra-access handover procedure and evaluation of its performance based on the user perceived performance aspects is discussed in [10]. The necessity of packet forwarding from a TCP throughput point of view and the problem of out of order packet delivery during handover are analyzed with a simple proposal. Finally, the impact of HARQ/ARQ state discard at handover on the radio efficiency is investigated. It is shown that due to a relocation based handover scheme (hard handover), the user perceived performance will not degrade. Restarting user plane protocols like RLC/MAC at the target cell is not making any radio inefficiency, however for achieving high TCP throughput performance, it is crucially important to employ packet forwarding and ensure the correct delivery of the packets. Also [11] investigates the performance of the handover procedure within LTE in terms of the delay of the whole HO procedure and HO failure rate. The results indicate that the handover procedure of LTE satisfies the goal of high performance mobility.

Computational complexity and Handover delay are two important factors to be optimized. Limited availability of line of sight (LOS) signals is one of the core reasons for call drop during handover in urban areas. Handover initiation in LTE is either based on RSRP or RSRQ, while the UE is measuring RSRP of the neighboring cell which is considered to be a candidate. If the RSRP of one of these candidates becomes better (larger) than in the serving cell, then the handover will be triggered. If handover decision cannot be obtained using RSRP, RSRQ will be used instead.

The impact of velocity, direction of movement, and propagation environment on handover performance of user equipment is analyzed by proposing a general model in [12]. Conventionally, handover is triggered using an instantaneous RSRP and RSRQ based procedure or a fixed averaging window. In the proposed method, an adaptive time window is given to provide additional flexibility for the handover procedure with the objective of decreasing call drop rate which (from user perspective) can be even more annoying than call blocking. In the proposed method in [12], instead of using conventional measures like RSRP and RSRQ, an average over the received signal has been made with the time window W which is self adaptive and a function of the UE velocity and the difference between the power received from the neighboring cell and serving cell rather than conventional measurements of RSRP and RSRQ, where the averaging window is fixed.

In [13], handover prediction is always considered as a method to increase the handover performance, although it is stated that the variety of proposals in this area increase system complexity, thus having less gains comparing to their cost. In [13], a prediction method is proposed, where a user mobility database for assisting the mobility prediction based on the user mobility history records is developed. The prediction of mobility and forecasting the future location of the UE for performing fast and seamless handover is considered as a matter of interest. The authors believe that the whole movement of the UE is not random. Especially in a cellular network, the information about UE movements can be used to optimize the handover algorithm. In the same area and in [14], a simple handover prediction based on acts of users of mobile history is proposed with the usage of user mobility database and simple mobility pattern matching. An approach for updating the database is also proposed, to ensure that is tracking the user mobility actions. Some methods for handover prediction are also proposed in [14], where a mobility management technique using handover prediction based on cross-layer architecture is discussed. The proposed prediction technique uses simple moving average and simple mobility pattern matching to decrease the total number of handovers in addition to the number of ping pong handovers.

In [15], a simulation analysis of handover performance in an LTE macro deployment with considering the impact and interdependency of handover parameters such as offset, TTT, filter coefficient K with different UE speeds and with the objective of reducing the handover failure rate by selecting the best combination is performed. It is concluded that different parameter combinations give the same HO failure rate assuming approximately the same mean time between handovers (MTBH). Using some methods that improve the reliability of control signaling can help to reduce handover failures.

A good setting for Time To Trigger (TTT) can mitigate ping pong handover [16]. Selecting too large values for TTT may cause undesirable radio link failure due to delay. Therefore optimal values for TTT that minimize the amount of ping pong handovers rate within allowable RLF rate can be found in [16]. Two methods, adaptive and grouping, are proposed and investigated to select the value of TTT. In the first method, the TTT is adaptively selected for each UE speed based on RLF rate of 2% while in the latter method the UE speed is classified into three different categories and the proper TTT is assigned based on what category the UE speed belongs to. The adaptive method significantly improves performance compared to the case when a fixed TTT value is applied.

A received signal strength based hard handover algorithm and its impact on the system performance, in terms of number of handovers, time between two consecutive handovers and the uplink SINR is investigated in [17].

In [18], link drops are proposed as another performance metric for the handover algorithm while the common metrics for handover performance evaluation are handover initiation delay and handover rate. The whole objective is again to provide the optimum combinations of handover parameters like hysteresis margin and threshold settings to improve performance metrics. The proposed metric (link drops) is found to be useful especially in finding the optimum operating point to minimize the outage in handover initiation.

A simple load balancing algorithm based on automatic adjustment of handover thresholds is proposed in [19] with the objective of reducing the call blocking rate and increasing cell-edge throughput in LTE networks.

Buffering the arriving packets in the network that may result in a sudden peak in utilization and buffer overflow is considered in [20]. The subsequent bursty packet losses will cause TCP connections to significantly degrade the throughput. Regarding this an analytical model for buffer overflows probability calculation, when the congestion window increase function and network parameters are given, is derived. An analytical method for calculating buffer overflow probabilities during mobile handovers is developed and this method can be used for any congestion control algorithm for which the elementary window increase function is known.

In [21], the femtoCell architecture of LTE is investigated and different handover scenarios are considered. [21] proposes two mobility management schemes at the Radio Network Layer (RNL) and for them complexity, signaling cost, standard impact and application scenarios are discussed.

In [22] and [23], the relation of handover and QoS is discussed. The first one deals with seamless HO and maintaining the quality of service (QoS) criteria with a fast HO decision algorithm. This fast HO decision algorithm is proposed for coping with the corner effect which is due to the loss of line of sight (LOS). The second one introduces an OFDMA handover algorithm which is known as the sub-carriers bidirectional arrayed handover algorithm (SBA). It is concluded in [23] that the SBA helps the users to maintain the best channel SIR and QoS, thus improving OFDMA system performance.

The usage of the X2 interface in forwarding the downlink user data between the involved eNBs is discussed in [24] and the impact of the forwarding on the user connections is investigated. In [25], the performance of TCP during LTE handover is discussed with a close look towards mobile users with high bit rates considering TCP services and the impacts of the intra LTE handover over their perceived throughput. Handover may cause a sudden degradation for the communication quality if the process is not correctly controlled. To decrease the effect of this degradation, three solutions are proposed in [25] which are categorized as fast path switch, handover prediction, and active queue management. The two first solutions are proposed for decreasing the delay and the third solution is active queue management which is designed for transport network. At the end it is shown that handover prediction together with active queue management can increase the performance of TCP significantly.

Data forwarding from the source eNB to the target eNB has also been considered in [25]. Data forwarding may cause an increase of delay in the packet delivery, especially when the network is congested. Delay may cause a retransmission timeout expiration of TCP and dropping of the congestion window. The TCP throughput of mobile equipment degrades significantly during the handover and that is due to forwarding of the packets from the source eNB to the target eNB. The forwarding avoidance algorithm consists of a fast path switch and prediction of handover. It is concluded that the prediction of handover is a very promising solution to avoid forwarding. The data forwarding approach that buffers user data (packets) when

Mobile Node (MN) makes a handover is investigated in [26]. Although this method promises seamless mobility, meaning no user data loss and short disruption of mobile node communications, when a large number of MNs make handovers simultaneously the required amount of buffer substantially increases. In [26] with the objective of the reducing buffered user data in BS while data forwarding, the relation between the amounts of buffered user data, the propagation delays of signaling messages and the user data are considered. Some modifications are proposed to decrease the amount of buffered user data. Simulation results show using the changes in handover procedure can reduce the amount of buffered user data to one-tenth of the amount usually required.

Seamless mobility in the networks that are using proxy mobile IPv6 is studied in [27]. In handover time the data is forwarded from the source Mobile Access Gateway (MAG) to the destination (MAG) and buffered there until a Mobile Terminal (MT) attaches to the new MAG. The algorithm proposed in [27] is to reduce the amount of buffered user data. The proposed scheme is based on discarding the forwarded user data in the case where the user data has large jitter, and discarding has no substantial impact on the quality of the application. In [28], three different data forwarding schemes for the networks that supports proxy mobile IPv6, are proposed to reduce the packet losses and the delay during handover.

A scheme with bicast and forwarding for handover (HO) between different base stations (BSs) is proposed to compensate for IP packet loss. The objective of the study is reducing the control delay for handover by transferring the IP packets both to the old and new BS before HO, i.e. cell change [29]. The proposed scheme also includes forwarding the IP packets from the source eNB to target eNB to provide lossless handover but only if they have been buffered before the start of bicast.

In [30] it is stated that by transmitting status information and forwarding buffered data in the interworking module, it is possible to recover frame loss occurring in handover quickly. The simulation models indicate a better performance because of quick data recovery due forwarding without loss of buffered data in case of handover.

The amount of bandwidth needed for X2 is discussed in [31]. ENBs are connected through X2 and in the time of handover, the data is forwarded via X2 from the source eNB to the target eNB. When only one user is making the handover, the bandwidth requirements for X2 for both control-plane traffic and user-plane traffic might not be so high, but when a group of users are crossing the cell borders and making handovers simultaneously, then the amount of required bandwidth might be high. However, when comparing with S1, the bandwidth requirement of S1 is much higher than X2.

4 Relays

4.1 Introduction

Deploying relays can overcome coverage limitations, especially at the cell-edge. Relays are mainly placed in coverage holes or cell-edges and can work as amplify and forward (AF) relays or as decode and forward (DF) relays [32]. Relays have the ability to increase absolute coverage as well as service coverage although they cannot increase the capacity since all the traffic should, nevertheless, pass through a macro. The connection of the relay to the macro is via a wireless backhaul, and on the other side, the macro connects the relay to the core network. In addition to the performance improvements the relays bring to the network, they can provide challenges in different areas.

In this chapter, the main functionality of relays is discussed first. After that different relay classifications are explained and finally relay system architecture is considered.

4.2 The main functionality of relays

In general, coverage as the basic requirement of any wireless network has been always the matter of interest for network operators. Extending the coverage, taking care of coverage holes, increasing the coverage of services, and providing high data rates in a larger portion of absolute coverage are the issues with high importance. Therefore providing solutions like deploying relays in LTE networks can facilitate the extension of coverage in new areas by improving the cell edge throughput. The relays are plug and play nodes with wireless backhaul, and installing and using them in the network should be much easier than regular eNBs. Thus using relays is cheaper comparing to eNBs which can be the dominant factor for their deployment. The ease of deploying relays makes them as an efficient temporary network deployment solution. The whole picture of the RN is shown in Figure 13. In this figure, the Uu interface is the access link and the Un interface is the backhaul link.

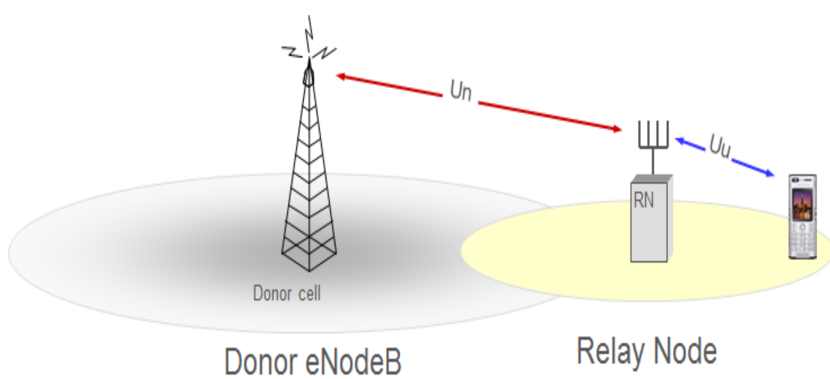

Figure 13: Relay served by a DeNB.

There has always been a great deal of concern about cell edges. The users who are located at cell edges cannot enjoy a good SINR and therefore cannot get high data rates. In order to provide high data rates for the users that are far away, eNBs should use a lot of power. However, using relays as intermediate nodes can help the operators to extend the coverage area of high data rates. Relays in their very basic form can amplify and forward the signal towards the user. In this case the received signal by the user will be more powerful. Also amplifying the signal for the users in the cell edge increases the cell edge throughput. Finally, relays can improve the functionality of group mobility. In group mobility, all the users that are served by a specific relay can be handed over to another node in the network. It should be noted that group mobility is not supported in Rel–10.

4.3 Relay classification

Relays can be categorized into different categories based on their radio perspective, frequency band usage, transparency, radio resource management strategy, and multiplexing scheme which are covered in the following.

4.3.1 Categorizing relays from radio perspective

From radio perspective relays can be divided into two main categories: layer 2 and layer 3 relays. The protocol stack of layer 2 relays in the second layer only comprises the MAC sub layer while the protocol stack of layer 3 relays in the second layer includes the MAC, RLC and PDCP sub layers. Both layer 2 and layer 3 relays can suppress the noise and the interference and they can also regenerate the signal. Although, on the down side, increased delay, resource splitting and heavy standard impact can be considered as their disadvantages. Comparing these two types of relays, layer 3 relays have less standard impact and better backward compatibility. There is also another type of relay which is called a layer 1 relay. Layer 1 relays are conventional repeaters.

4.3.2 Inband and outband relays

Considering the relay node spectrum usage, relay is categorized into two classes of inband and outband relays. In inband relaying, the backhaul link (Donor eNB-Relay link) and access link (Relay-eNB link) use (share) the same frequency band, while in outband relaying the operation of Relay-UE link and Donor-Relay links are in different carrier frequencies. Inband relays are more interesting for operators because the operators are not necessarily willing to reserve extra carriers for relays due to the high costs and bandwidth limitations.

4.3.3 Transparent and nontransparent relays

The third classification of relays is regarding to the UE's awareness about the existance of relays. There are two classifications of transparent and nontransparent relays. In transparent relaying, the UE is not aware of the existence of the relay if it is communicating with the network via a relay, but in nontransparent relaying the UE knows that it is connected to the relay when communicates with the network via the relay.

4.3.4 Relay classification based on strategy

The forth classification concerns relay strategy to be the part of the Donor cell or to have control over its own cell. If the relay is part of the Donor Cell the relay will not have a cell identity for its own while it might have a relay ID. In this case part of the radio resource management (RRM) is controlled by the Donor eNB and the rest is controlled by the relay. Smart repeaters, decode and forward relays, and different type of L2 relays can be examples that use this strategy.

In the other strategy the relay controls its own cell or cells. In this scenario, in a unique cell an identity will be provided for each cell, the same RRM mechanism as the eNB will be applicable for relays, and the UE will not see any difference in connecting to the relay or the eNB. For instance L3, relays make use of this strategy.

4.3.5 Type 1 and type 2

Type 1 relay controls its own cells, and the cells have their own physical cell ID. Type 1 relay node transmits its own synchronization channels and reference symbol. In single-cell operation the UE receives scheduling information and HARQ feedback directly from the relay node and sends its control channels to the relay node. There are also type 1a and type 1b relays. These two relay types have the same characteristics as relay Type 1. The only difference between these two relays is that relay Type 1a is the operating outband while relay Type 1b is the operating inband with sufficient antenna isolation.

The type 2 relay does not have a separate cell ID, therefore this type of relay will not create new cells. The type 2 relay has the possibility of transmitting PDSCH.

4.3.6 Relay separation by Un /Uu transmission

Relays can be separated based on the ability for simultaneous transmission support in Un and Uu. If the relay is outband or there is sufficient isolation between access and backhaul antennas, there can be simultaneous transmission in Un and Uu links. But in case of inband relays without sufficient isolation between access and backhaul antennas, using time multiplex between access and backhaul provides separation between Un and Uu transmission. This is necessary to avoid interference issues.

4.4 Relay system architecture

In the LTE network deploying relays, from a protocol termination perspective, the relay as an intermediate node terminates all the radio protocol layers towards the UEs. The relay also terminates the S1 and X2 protocols towards the rest of the network.

In the adopted architecture for Rel-10 LTE relays, S1 and X2 proxy functionality is provided by the DeNB. The RN is considered as new cell under DeNB and identifies the DeNB as the MME in the S1 protocol and as the eNB in the X2 protocol. The DeNB acts as a serving PDN-GW for the relays that creates a session for them and manages EPS bearers for the relays.

4.4.1 Relay-eNodeB link for inband relay

In the LTE network deploying relays, the backhaul link is called the Un link and the access link is called the Uu link. In inband relaying, the transmission in these two links cannot happen simultaneously, therefore the resources to be used in these two links are separated in the time-frequency space which is known as resource partitioning. In this structure in the downlink the eNB-RN and RN-UE links are using time division multiplexing in a single carrier frequency that means only one of the links can be active at a time. In the uplink RN-eNB and UE-RN, it is the same and only one of the links can be active at a time.

In inband relaying self-interference is an important issue where the relay transmitter causes interference to its own receiver, and therefore simultaneous eNB-RN and RN-UE transmissions on the same frequency resource both in uplink and downlink are not possible unless there exists a sufficient isolation of transmitting and receiving signals. One solution to overcome this interference problem is to make time gaps so that the UE does not expect any transmission from the relay at the time the relay is receiving from the Donor eNB. The gaps where the UE should not expect any relay transmission are made by MBSFN subframes configuration. In MBSFN subframes the UE does not expect downlink transmission from the relay. Uplink subframes follow downlink subframes.

In inband relaying the backhaul link can use FDD or TDD multiplexing. In FDD multiplexing, the eNB-RN transmissions in the uplink and downlink are done at different frequencies which are dedicated for uplink and downlink transmissions. In TDD multiplexing, the eNB-RN transmissions in the uplink and downlink are done in different subframes.

4.4.2 Relay-eNB link for outband relay

If the access and the backhaul links have enough frequency isolation for example with the help of antenna separation, then there is a possibility of simultaneous transmission and there will be no interference issue happening due to simultaneous transmissions. In this scenario there is no need for time division multiplexing.

4.5 Former research

A large amount of research has been done on relays to find solutions for the challenges. Some of the most recent studies on relays are covered in the following as examples:

In [33], a system-level simulation has been performed to evaluate the performance of Type 1 relay with realistic model for backhaul link and the consideration on the limitations to access link scheduling. Further work has been done by studying the backhaul link as a limiting factor in relay deployment as well as a comparison of the relays with pico nodes and traditional homogenous macro cells [34]. A relation between RN transmission power, a ratio between the number of Base Stations (BSs) and RNs, and the performance of the system is given in [35]. Indoor coverage extension by using Amplify and Forward (AF) or Decode and Forward (DF) relays is studied in [36]. RRM strategies for Type-1 relay nodes are investigated in [37] considering the backhaul link as the bottleneck. In [38], the advanced adaptive resource sharing strategy for efficient resource utilization among relays in a cell is studied. Finally some of the handover challenges in LTE networks deploying relays are covered with possible solutions in [39] [40] [41] [42].

5 Problem formulation

5.1 Introduction and problem definition

The handover procedure is discussed in Chapter 3. As it is depicted in Figure 9 one part of the handover procedure is data forwarding. By data forwarding we mean sending the data from the source to the target at handover time in order to avoid losing any data or transmitting data twice. In Chapter 3, the former research on data forwarding and the amount of gain that data forwarding can provide are discussed in LTE networks. In the same chapter, the research on handover procedure in LTE networks containing relays is introduced. The research can be mostly concluded as the changes in handover procedure to provide a more reliable handover. In LTE networks with relay deployment, research in different areas has been conducted as is shown in Chapters 3 and 4; however no research is reported regarding data forwarding in LTE networks containing relays.

Relays as new nodes in the network bring their own challenges. After proposing the idea of using relays in the LTE networks, a huge amount of research has been done in different areas related to the relays to verify their effects in the network, proposing solutions for the challenges and finding out the performance of the proposals. As the mobility has been always a challenge in the networks, after proposing relays in LTE, mobility should be studied carefully. One of the mobility challenges that deploying relays in LTE networks introduces is in data forwarding scheme in the handover execution time. The problem occurs when the UE which is served by a relay makes a handover. To have a lossless handover, in handover time the data which is in relay should be transferred to the target and from the target to the UE.

In normal downlink data transmission in a relaying scenario, data is sent from the Donor eNodeB to the RN and from the RN to the UE. In handover time, the data that is not acknowledged by the UE should be retransmitted back to the Donor eNodeB from the RN and then from the Donor eNodeB it should be sent towards the target. In Figure 14, this is depicted. It can be seen that there is some data that should be sent between the Donor eNodeB and RN back and forth. When the UE detaches from the serving relay the data in relay which is unacknowledged by the UE should be retransmitted to the Donor eNodeB. In addition some amount of ongoing data from the Donor eNodeB towards the UE cannot be received by the UE. Thus this amount of data should be transmitted back to the Donor eNB to be transmitted towards the target. Data transmission from the Donor eNodeB towards the RN and again back towards the Donor eNobeB is called the back and forth data forwarding problem. Back and forth data forwarding is clearly a waste of Un resources. As the data to be forwarded should be transmitted in uplink in a wireless backhaul from the RN to the Donor, it causes extra delay in the system especially because of the need for uplink transmission. This delay decreases the bit rate received by the UE. To conclude the two main problems that back and forth data forwarding brings up are (i) wasting Un resources and (ii) causing additional delay.

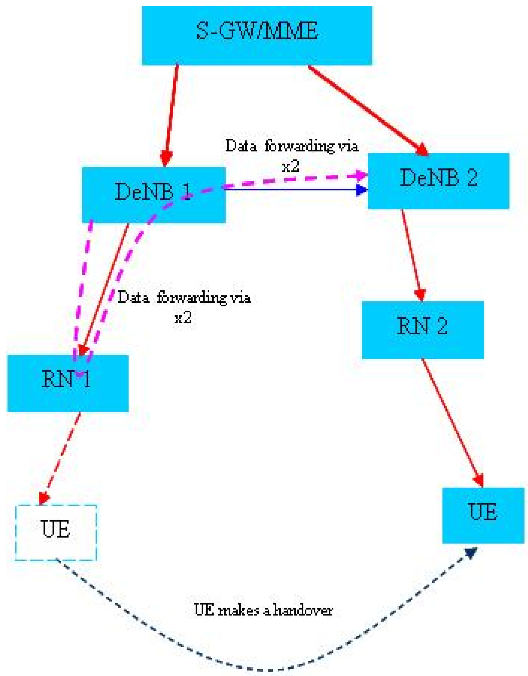

Figure 14: Back and forth data forwarding.

Back and forth forwarding has been noticed and solutions to fix this problem have been proposed. However, no accurate studies using simulations have been done so far and no publications have been reported in this regard.

In the rest of this chapter first we take a closer look at the back and forth data forwarding problem and then discuss different solutions to solve the problem mentioning their pros and cons.

5.2 Different handover cases

Having relays in LTE, different handover situations might happen based on to which node the UE is connected and to which node it is handed over. In total six handover cases can be possible that are classified in the following.

1. handover from the RN to the RN when the relays are served by the same eNodeB
2. handover from the RN to the RN when the relays are served by different eNodeBs
3. handover from the RN to the eNB which is serving the same RN
4. handover from the RN to the eNB which is different from DeNodeB of the RN

5. handover from the eNB to the RN which is served by that eNodeB
6. handover from the eNB to the RN which is served by another eNB.

As it was mentioned before, the back and forth forwarding problem only happens if the UE is connected to the RN before handover; therefore the cases 1 to 4 mentioned above are handover cases where the problem can occur.

Case 2 was depicted in Figure 14. Cases 1, 3, and 4 are shown in Figures 15, 16 and 17. The Back and forth forwarding problem is shown in these figures by a dashed arrow. It can be seen, the transmitted data from the Donor eNB to the RN should be transmitted back to the Donor eNB.

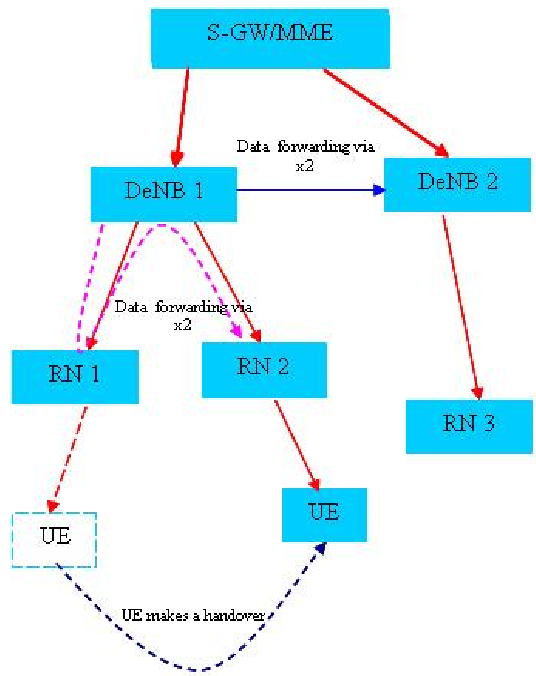

Figure 15: Data forwarding in handover case 1.

5.3 Motivation for the study

One of the problems that the back and forth data forwarding brings to the handover procedure is the extra delay which decreases the system performance in general. Therefore finding a solution to decrease or to remove the extra delay to optimize the performance is inevitable. As it was perceived from the former studies in Chapter 3, one of the main factors that can improve handover performance is to have as short an interruption time in the user plane connectivity as possible. The back and forth data forwarding in relay scenarios prolongs interruption time. Especially when the network is more congested, it takes time for the packets to be transferred from the source to the destination since the packets are to be queued in some nodes.

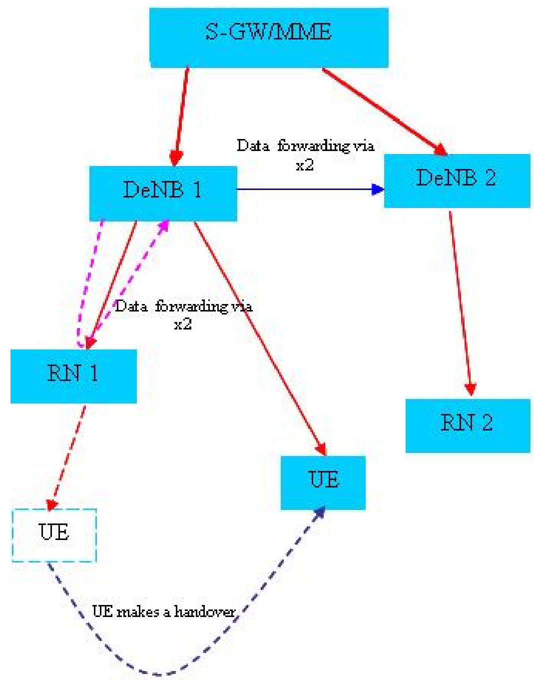

Figure 16: Data forwarding in handover case 3.

In the networks without RN deployment, in handover time the data in source eNB is forwarded to the target eNB via an X2 link, but in networks with relay deployment, when the UE is connected to RN, the packets should be first transmitted back in the uplink of the wireless backhaul link between the RN and the Donor eNodeB which is considered as the bottleneck since the bit rate in the uplink is less than the downlink and it causes more delays. The second problem is using some of the available resources in the Un link and occupying them unnecessarily. Resources are rare and valuable, so whenever there is a waste of resources optimization is inevitable.

Understanding the problem can lead us to the general idea for solving the problem. The amount of data that is transmitted back and forth should be completely avoided or should be as little as possible. For instance, dropping all the data that should be transmitted in data forwarding can totally solve the problem of back and forth forwarding although, on the other hand, it puts the whole data forwarding procedure under a question mark, and may affect the downloading bitrate of the user to some extent which might not be desired. In the rest of this chapter some of the possible solutions are discussed and compared with each other.

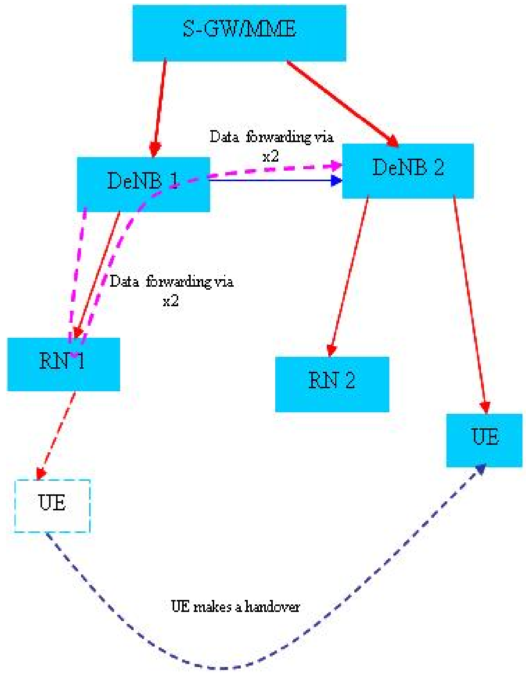

Figure 17: Data forwarding in handover case 4.

5.4 Possible solutions

5.4.1 Buffering PDCP SDUs in DeNB

In 3GPP contribution [43], a solution where the DeNB keeps (buffers) the PDCP SDUs until they are acknowledged by the UE is mentioned. This idea is not discussed further in the contribution but we investigate it here. By a closer look we can notice that this idea can solve the problem of back and forth forwarding. The root of the problem can be considered as unawareness of Donor eNodeB about the UE status.

If the DeNB knows about the behavior of the UE, especially if it knows the packets that have not been acknowledged by the UE, the back and forth forwarding problem will be solved. If the Donor eNodeB would keep the PDCP SDUs untill they are acknowledged by the UE and then would delete them, if in handover interruption time some of the packets are not received by the UE, the Donor eNodeB can simply forward those packets to the target eNB to be transmitted towards the UE as soon as the UE attaches to the target eNB. This approach can solve the problem but raises the question of how the Donor eNodeB can be aware of the UE behavior while the UE is connected to an RN.

In [44], the same idea has been proposed in addition to the possible ways that the Donor eNodeB can be aware of the UE status. As discussed above, the DeNB can forward the packets that are not acknowledged by the UE towards the target eNB. In RLC-AM, the DeNB could discard the PDCP SDUs only when they have been received by the UE and in RLC-UM when they are sent to the UE. For letting the DeNB know about the UE and the existence of a capability of ACK/NACK between the UE and the Donor eNB:

1. The RN can send a periodic PDCP status report to the DeNB and inform it about the PDCP PDUs that are received by the UE [44].
2. An associated PDCP SN in the Un and Uu links can be used. In the current Relay design, the protocols are different for Uu and Un interfaces. As an example, if in Un link which is the interface between the DeNB and the RN, we have PDCP SN1 for a given packet, whereas, in the Uu link which is the interface between the UE and RN we have PDCP SN2. These two PDCP sequence numbers could be associated with each other [44].

5.4.2 Handover anticipation

The second idea to solve back and forth forwarding is to forecast the happening of the handover. Then the DeNB can stop sending the packets towards the RN based on the anticipation [44]. The accuracy of the anticipation is very important. If the handover happens sooner than the forecast back and forth data forwarding problem occurs. Happening of the handover shortly before the forecast can overcome back and forth forwarding problem to some extend. If the handover happens later than the forecast, some of the packets are not delivered to the UE, because transmission to the RN from the DeNB has already stopped. These packets will be forwarded to the target cell and will be delivered to the UE when it attaches to the target cell.

5.4.3 Buffering downlink data upon handover request reception

In the third solution proposed in [45] once the DeNB receives a handover request from the RN, it stops transmitting the data to the RN and starts buffering the downlink data. This idea will make a small change to the handover procedure. As the DeNB receives the handover request from the RN it stops transmitting the data to the RN and starts buffering the UE-related downlink data. After that, the Donor eNodeB sends the acknowledgment to the RN handover request and the normal handover procedure can start. Then the packets those have been already sent to the UE and have not been acknowledged yet will be forwarded towards the Donor eNodeB, possibly with SN status, and the DeNodeB will reorder these packets, the packets that are buffered and also the packets that have come from the Serving Gateway and sends them all to the target cell [45].

Although it is claimed in [45] that the back and forth forwarding problem is completely solved, the packets that are in the RN at handover time, in any case should be transmitted to the target eNodeB via the Donor eNodeB. It is obvious that these packets have reached the RN from the Donor and now should be retransmitted again to the Donor eNodeB. Therefore the back and forth forwarding problem still exists, but this idea decreases the amount of packets that are going back and forth.

5.4.4 Buffering downlink data upon handover request reception as well as transmitting them towards RN

In the fourth solution also presented in [45], the Donor eNodeB does not stop transmitting the packets towards the relay once it receives the handover request. This method is a combination of buffering the packets and transmitting them at the same time. As the Donor eNodeB receives the handover request which is sent by the RN, it sends a starting marker to the RN to be the reference for synchronization and then starts to buffer the data and sends the data at the same time [45]. Sending the start marker is necessary because this marker is considered as a starting point; therefore it can be used as a reference for the packets that are received in the RN and the packets that are buffered in the Donor eNodeB to follow the same numbering regarding to this reference. Then at the time that the UE detaches from the RN, the relay can send the status report to the DeNodeB and inform it about the packets that are acknowledged by the UE as well as the packets that are not acknowledged by the UE. So the DeNodeB can find those unacknowledged packets from its buffer and send them to the target eNodeB.

In this method there will not be any need to send the forwarded packets in Un. Therefore the problem of back and forth data forwarding will be completely solved except for the packets that in downlink Un are on the air on their way to the UE in detaching time. It is obvious that these packets cannot be acknowledged by the UE and they should be retransmitted back to the Donor eNodeB. This can be true for the third idea as well. In the forth idea there is a need for sending the status report in the uplink from the RN to the target eNodeB that uses the Un resources. The other issue is buffering the data in the Donor eNodeB. If many users want to perform handover at the same time, this may cause to buffering of many packets and larger buffers are needed.

5.4.5 Discussions

In the first solution, there should be a one to one mapping of sequence numbers of PDCP PDUs on Un and Uu interfaces. The RN sends a PDCP status report to the DeNB to give the DeNB the information about the UE status. Applying this solution, the DeNB can be aware of the packets that are received by the UE and the packets that are not received by the UE. In the second solution, the DeNB should stop transmitting to the RN as soon as receiving the handover request. Using these two solutions can diminish or totally remove back and forth data forwarding.

Comparing the third and the forth solutions; in the third solution the Un resources are not used at all or are used in some extent comparing to the default implementation (all the data both in the DeNB in the detach time should be transmitted to the RN and then from the RN back to the eNB). There is no need for sending start marker or end marker, because there will not be any need for synchronization for the transmitted and buffered packets. The drawback of the third solution might be the amount of delay that it provides. While if we consider the forth solution, because more packets are transmitted to the UE from the DeNB before the UE detaches, it is more probable to have less delay. Two drawbacks can be mentioned for the fourth approach; one of them is sending the start marker to start the synchronization and sending the end marker to show the end of transmission as well as sending a handover status message. The other disadvantage is wasting the Un resources.

The necessity of using an end marker is emphasized in [46]. In the path switch, the serving gateway sends one or more “end markers” in the old path to show the source eNB that no more data will be sent in that path. On the other hand, when the source eNB forwards the data to the target eNB, the end markers can show the end of data forwarding. But when an Intra-DeNB handover happens (the source RN and the target RN are both served by the same DeNB), then the MME and serving gateway cannot find any change in the path; because the serving eNB has not been changed. Therefore the serving gateway cannot send any PATH SWITCH and “end marker” commands. Then the DeNB and RN cannot notice the end of downlink as well as data forwarding. To solve this problem, the DeNB can send an end marker to the RN in the Un link. The end marker sent by the Donor eNodeB has exactly the same format as the end marker sent by the serving gateway. As mentioned above for the intra-DeNB handover, the DeNB and the RN cannot notice that there is going to be a path switch from the CN, but when the DeNB receives the handover request from the RN, it can forecast an upcoming path switch which should be done by the DeNB itself. Therefore when the DeNB receives a handover request, it sends an end marker to the RN and then the DeNB changes the path to the target eNB. It is obvious that the path between the serving gateway and the eNB will not be changed, and even after the handover, the serving gateway will not know about the handover occurrence.

The same solution as [45], is proposed in [47] with the same approach. Two well known problems that occur, redundant data forwarding first from the DeNB to the RN and again from the RN to the DeNB, are addressed. As it was mentioned earlier these problems increase the user data latency and wasting of radio resources. Some resources would be wasted due to the unnecessary uplink and downlink transmission to forward the data. RN power is small that is why the probability of handovers will be very high. More handovers causes more data to be forwarded and therefore there will be additional waste of Un resources. Therefore an idea to avoid the back and forth forwarding problem is proposed. The DeNB sends a “Start marker for Relay”, SM-R, and then starts buffering the data as well as transmitting the data to the RN as it receives the handover request. Then when the HO happens, the RN sends an HO status report to inform the DeNB about the packets that are received by the UE and the ones that are not received. In addition, in [47] an structure for the HO status report is proposed, which has the SDU’s sequence number which no

SDUs have been acknowledged after that.

5.4.6 The amount of data going back and forth in Un link

The amount of data forwarding in Un link is estimated in [48] and [49] using exactly the same approach. The authors first try to show the amount of data which is forwarded in the Un link. The amount of forwarded data can be calculated by multiplying the data rate by the time duration that the data is forwarded. The average data rate in the Un and Uu links can be assumed to be similar because the usage of flow control at the application level. Therefore there is a possibility to estimate the data rate in Uu using the cell edge data rate and from that reach the data rate of Un. The data rate at the cell edge is very low. By the assumption of the spectral efficiency of 0.06 bps/Hz/user, 10 active users in the cell and a 10 MHz system, the average data rate will be 600 Kbps. It is assumed that the time period between the moment when the DeNB receives the handover request ack until the path switch is around 81.3 msec. From these values the amount of data which is data rate $*$ the time period can be calculated as $600Kbps * 81.3msec = 48780$ bits.

For the whole cell, the spectral efficiency is assumed to be 0.19 bps/Hz/User, therefore the whole amount of data in transmission will be $0.19 * 10 * 10^6 * 10 = 19 * 10^6$ bits. Dividing these two values the data percentage which is transmitted in Un in handover time compared to the whole data transmission in Un can be reached which is less than 1%. Based on these assumptions and culculations the authors in [48] and [49] believe that the amount of data during handover is small time compared to the whole amount of transmission over Un and does not occupy and waste a lot of resources in the Un link. For the above calculation average values with simple assumtions are applied however, the amount of forwarded data can vary a lot based on many factors such as buffer sizes, AQM existance, relay positions, network congestion and if the Un or Uu links are bottlenecks. Considering these situations and doing quick simulations for calculating the values of forwarded data in the simlator we are convinced that finding a solution for back and forth forwarding problem is necessary otherwise Un resources will be wasted and some users will suffer from long forwarding delays.

Looking closer at [48], the same calculation has been done with slightly different values. The authors have assumed again the same system with 10 MHz bandwidth and 10 active users. Moreover, the authors have assumed the same data rate in Uu and Un and they have tried to calculate the percentage of the data in Un link in the data forwarding period compared to the whole amount of data in the cell. They have assumed an average spectral efficiency of 0.24 bps/Hz/user and average cell edge throughput of 0.07bps/Hz/user. The amount of time that they have defined as reduction time here is from the moment that the DeNB receives the handover ACK untill the path switch. It is assumed this time is around 100 msec. Then the amount of data in the Un link is 70 Kbits and the amount of data for the whole cell is 2.4 Mbits for 1 sec. The devision of these amounts will be equal to 2.9 percent which is considered as a low amount and the authors have concluded that the resource cost of redundant forwarding packets is relatively small. About the latency which is the other problem that the data forwarding causes the authors believes that handover optimization does not provide us with any gain when compared to the increase in cost and complexity [48].

6 Simulation results and analysis

6.1 Overview

Performance evaluation in the networks comprising multiple cells is a complicated task. Therefore creating an artificial environment similar to the real one with all the elements and protocols can be used as an alternative for performance evaluation. This environment is called a simulator and creating different scenarios in the simulator and running them is called simulation. Running simulations is not only a reliable way for different evaluations before performing implementation in real networks, but also a very good way to avoid complex calculations.

In the simulators a network including multiple cells is created, furthermore logical nodes in the network with their necessary protocols are implemented. If the simulator is a dynamic simulator, a UE arrives and moves in the network or if the simulator is a static one, the UE is dropped into the network randomly several times to cover all parts of the network. In dynamic simulators the UE moves in the whole network through a random or predefined path, does measurements, downloads files, faces fast and shadow fading, suffers from interference and, whenever it is needed, it makes a handover. In the simulator, the behavior of different nodes can be logged and viewed in a way similar to real networks.

In simulators, different layers with their protocols are implemented. For each layer specific elements in the protocol are defined to be viewed with logs. For example in the MAC layer, scheduler behavior can be shown or at the TCP level, congestion window behavior can be viewed. The possibility of having logs in different layers, not only helps us compare the performance of different approaches but also helps us to find the probable problems in the other network parts. As an example having some droppings in the TCP congestion window graph shows that we have potentially packet losses in the radio link.

The objective of running simulations is to verify and compare different solution proposals for the problem. After considering different approaches it can be concluded whether any of the proposed approaches works better than the others and better than the default one. In the next step, if conclusions show that the relation of gain versus complexity and cost is high enough, the selected approach can be implemented in real networks, potentially after standardization. In standardization, different companies bring their own proposals and after negotiations one of the proposals that convinces the others is selected.

The simulator used in this thesis is a dynamic full protocol simulator. Using this simulator, the handover performance in LTE networks containing relays can be studied. The back and forth forwarding problem is investigated and the different proposals made in Chapter 5 are simulated and compared. The rest of this chapter is organized as follows: first we explain the simulation model in Section 6.2. The next two sections cover the different approaches used for the simulations and performance metrics consequently. Finally, performance evaluation and comparison between different models is made.

6.2 Simulation model

The simulation is done in a network having 3 Donor eNodeBs and 9 relays. Therefore we have 12 cells in total. The network is a wrap around network. In a wrap around network, when the UE goes out of the network from one side, it enters the network from another side. The network is a hexagonal grid and for each sector of the DeNB, there is one relay. Similar to DeNBs, relays also have 3 sectors in the model. In Figure 18, the network is depicted. It is demonstrated in the figure that both relays and DeNBs have 3 sectors. The simulations are performed having one UE in the network. In simulation time, the UE arrives at the network in a random position, starts to download a file and then moves randomly in the network with a defined speed. The UE is removed as it finishes downloading the file. After that, the next UE arrives at the network and starts to download exactly after the former UE has disappeared. Each UE may experience one or several handovers while existing in the network and the number of handovers can affect the performance of the file download. This is investigated later in this chapter. The used handover model is detailed such that all the measurements and delays are implemented. The selected traffic model is TCP/IP FTP. The scenarios are used for downloads. Cell radius is considered as 166 meters and the eNB and the RN power are 20 Watts in downlink. Table 1 shows the important simulation parameters.

Table 1: Simulation parameters

Parameter	Value
Cellular layout	12 sites, 3 sectors per site
Cell radius	166 m
Propagation model	Typical urban
Carrier frequency	2 GHz
System bandwidth	5 MHz
eNB transmit power	20 Watts
RN transmit power in downlink	20 Watts
RN transmit power in uplink	0.25 Watts
UE transmit power	0.25 Watts
User drop	Random

6.3 Examined scenarios

To investigate how much deduction in latency and usage of Un resources in data forwarding can be gained, we evaluate the ideas proposed in Chapter 5 and calculate

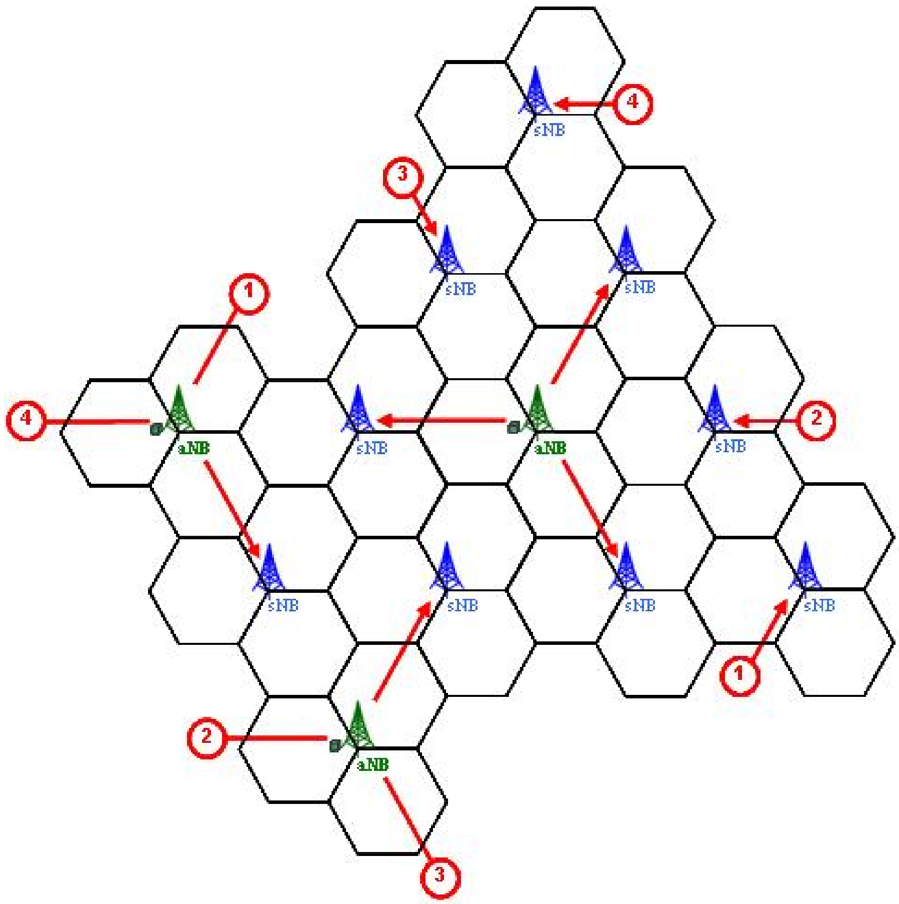

Figure 18: Network Model.

and compare the amount of forwarded data and latency in data forwarding for different approaches. As explained before, the forwarded data can increase the latency and consumes Un resources unnecessarily, which is not desired. To study the effects of data forwarding and to verify the solutions, four different approaches are developed. These approaches can be categorized in three divisions. The first category is called "default approach", demonstrating the existing way of data forwarding. The second one demonstrates the ideas for changing and improving the default model that are named as "optimized" and "Dropping PDCP SDUs" approaches. The last approach can be used as a reference since it is an idealistic implementation of data forwarding that does not cause any delay although all the data receives the destination. In the following different approaches are described in detail.

6.3.1 Default approach

As it was mentioned earlier, the back and forth forwarding problem exists only if the UE is connected to the RN. At the time UE detaches from the RN there are some packets in the RN that have been sent towards the UE but the acknowledgements for them have not been received by the RN yet. These packets therefore should be forwarded in the uplink towards the DeNB and from the DeNB to the target destination. Also new packets arrive from the S–GW to the RN through the DeNB until the path switch. It is obvious that recent packets cannot be transmitted to the UE from the RN at all since the UE has been already detached from the RN. Therefore these packets should be transmitted in the uplink towards the DeNB and then to the target destination. Back and forth forwarding continues until the path switch. The amount of data going back and forth has a direct relation to the time the UE detaches from the relay until path switch occurrence.

6.3.2 Optimized approach

In the second approach, which is called the optimized approach, as the UE detaches from the RN, the new packets from the gateway to the UE will be forwarded directly towards the destination as they reach the DeNB. Unlike the default approach, these packets are forwarded towards the destination without being transmitted through the Un link. The other packets that are in the RN buffer and have not been transmitted to the UE or have not received the acknowledgement from the UE yet, should be transmitted in the uplink to the DeNB and from the DeNB to the destination.

6.3.3 Dropping PDCP SDUs approach

The third approach which is called Dropping PDCP SDUs is similar to the optimized approach except for the packets that are in the RN buffer. In the droppoing PDCP SDUs approach, all the new packets originating from the gateway will be forwarded directly towards the destination from the DeNB in the detach time but the other packets that are in the RN buffer will be dropped instead of being forwarded.

6.3.4 Ideal approach

In the fourth approach all the packets will reach the destination without any packet drops and without any delay. In the ideal realization there is no delay between the UE's detaching from the old cell and attaching to the new cell. Path switch is also done immediately.

6.3.5 The amount of forwarded data going through Un link

In the default approach the amount of data which is involved in the forwarding process is transmitted back and forth through the Un link. Back and forth data transmission while data forwarding creates large delays and occupies the Un resources unnecessarily. In this section we investigate the amount of data that is transmitted back and forth in the Un link using optimized approach and default approach.

In this section, the amount of forwarded data in the default and optimized methods for two different file sizes and two different speeds are shown. In the optimized approach, only the buffered packets in the RN in interruption time are transmitted in the backhaul uplink. In the default approach besides these packets, also the new packets from the gateway towards the RN are transmitted back and forth in the backhaul. These new packets in the default model pass through the backhaul and in the optimized model are redirected to the target without being sent in the backhaul as it was discussed in Chapters 5 and 6.

Figure 19 shows the amount of forwarded data using the optimized and the default methods for the file size of 1 MB and speed of 50 km/h. From the whole amount of forwarded data, 92.5% is belonged to the new data from the gateway to the RN and 7.5% is belonged to the forwarded data from the RN to the DeNB in the uplink. Therefore using the optimized model can decrease 92.5% of the whole amount of data with the forwarding need to go through the backhaul. The whole amount of forwarded data in this case is calculated to be 22.9% of the whole data file of 1 MB. Therefore using the optimized method, we can avoid 21.2% of the whole data file goes back and forth in the Un link.

In Figure 20, the file size is considered as 1 MB again but the speed is increased to 120 km/h. The figure shows the amount of forwarded data for optimized and default methods. In this scenario, 19.9% of the whole file size needs to be forwarded. 17.2% is related to the new packets from the gateway to the RN and 2.7% is belonged to the buffered packets in the RN. At this speed by using the optimized approach the amount of forwarded data can be reduced by 86.4%.

In Figure 21, the amount of forwarded data for the speed of 50 km/h and the file size of 40 MB is depicted. In this scenario the new packets from the gateway to the RN consist of 91% of the whole forwarded data and the buffered data in the RN consists of 9% of the whole forwarded data. In 50 km/h downloading the file with the size of 40 MB, the optimized approach avoids 91% of the whole forwarded data to be transmitted back and forth into the backhaul.

In Figure 22, we have the file size of 40 MB and the speed of 120 km/h. The whole forwarded data consists of the new packets from the gateway to the relay with the amount of 85% of the whole forwarded data and buffered data in the RN with the amount of 15% of the whole forwarded data. Using the optimized approach in this scenario, 85% of the forwarded data is avoided to pass through the backhaul.

In all the cases more than 80% of data with the need of forwarding can be reduced to be transmitted back and forth in the backhaul. Reducing the amount of forwarded data can have a large effect on reducing the use of backhaul resources and delay. Also it can be seen that when increasing the speed, the amount of new packets coming from S1 interface decreases. This is because with higher speeds, the bitrates of the radio links are lower. The TCP congestion window limits the source rate to match the underlying link speed. With lower source rates, also buffer sizes are smaller.

In the next subsection we consider the four scenarios again and we look into how much delay reduction we can obtain by using the optimized method.

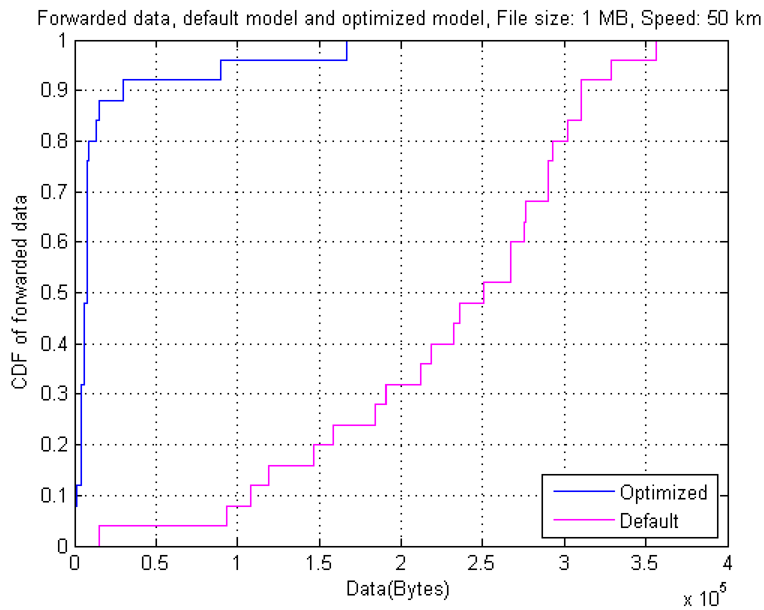

Figure 19: CDF for forwarded data with default and optimized approaches, file size: 1 MB, speed: 50 km/h.

6.3.6 Interruption time in the user plane

Forwarding of a large amount of packets can increase the delay in the whole process of file download. Especially transmitting packets in the uplink adds to the amount of delay. Applying the optimized approach can be a reasonable way for decreasing the amount of forwarding delay. In this section we show how much delay can be decreased by using the optimized approach rather than the default approach.

The interruption time in the user plane is the consequence of data forwarding. The target eNB should wait for the forwarded data to be received before it can transmit the forwarded data or the new data to the UE. Uplink is considered as the bottleneck for data transmission, thus data transmittion in the uplink takes more time than downlink in general. Sending a large amount of data back and forth in the backhaul results large latency. Also if the uplink is congested, the packets stay in a queue before being served that increases the latency. In the former subsection it was shown, using optimized approach can decrease the amount of packets to be transmitted back and forth, therefore we expect, the optimized approach can decrease the interruption time in the user plane.

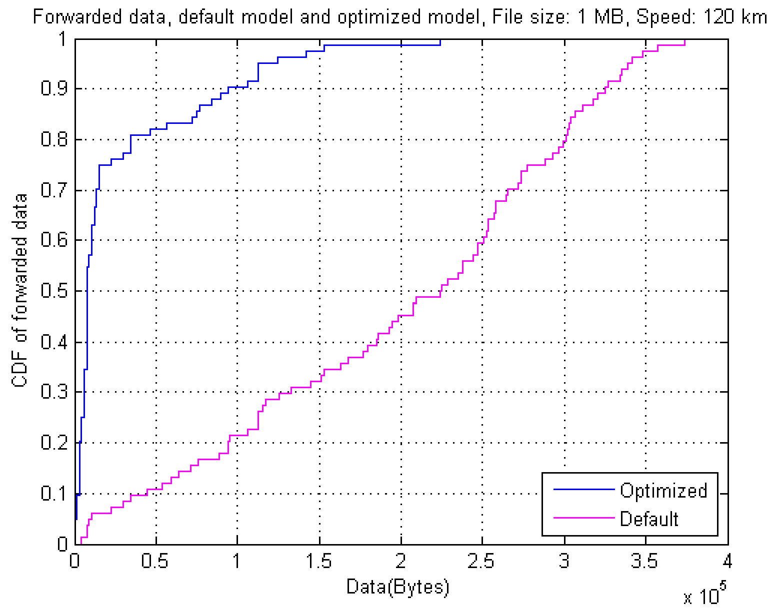

Figure 20: CDF for forwarded data with default and optimized approaches, file size: 1 MB, speed: 120 km/h.

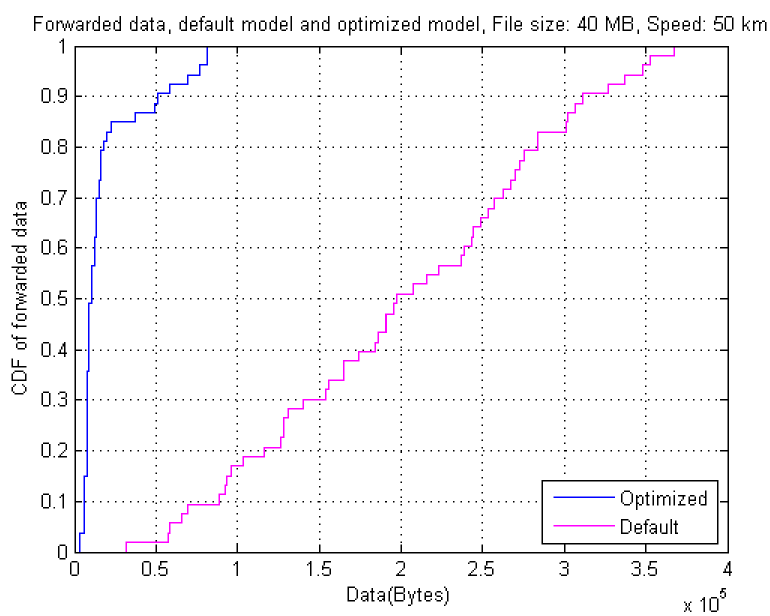

Figure 21: CDF for forwarded data with default and optimized approaches, file size: 40 MB, speed: 50 km/h.

In this section, four scenarios similar to subsection 6.3.5 are considered. In Figure 23, the interruption time in the user plane, when the speed of the UE is 50 km/h and the UE is downloading a file with the size of 1 MB, for optimized and default approaches is shown. In this scenario, comparing the optimized and default approaches, we measured 94% delay deduction. Figure 24 shows the interruption time in the user

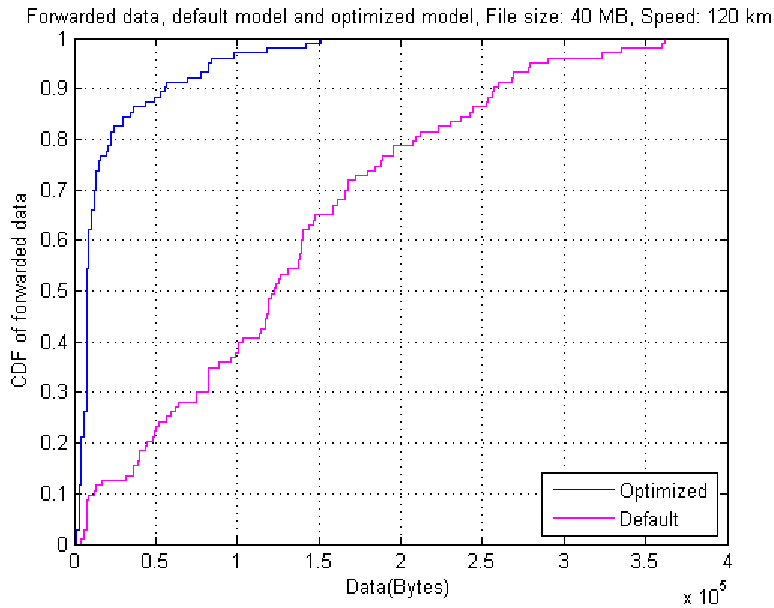

Figure 22: CDF for forwarded data with default and optimized approaches, file size: 40 MB, speed: 120 km/h.

plane for the speed of 120 km/h and file size of 1 MB. In this scenario optimized approach shows 91% delay reduction comparing to the default approach. In Figure 25, the interruption time in the user plane for the speed of 50 km/h and file size of 40 MB is depicted. In this scenario using the optimized approach can deduce the delay by 89% comparing to the default approach. In Figure 26, the interruption time for speed of 120 km/h and file size of 40 MB is illustrated. In this scenario, the optimized approach can reduce the delay by 83% compared to the default approach.

As a conclusion, it is seen for all the scenarios, applying the optimized approach instead of the default approach, the delay reduction is more than 80%. Also by an increase in the speed the portion of the delay which is related to the forwarding of new packets coming in S1 interface decreases. This is because by increasing the speed, the amount of new packets coming in S1 interface as explained in 6.3.5 decreases.

6.3.7 Handover classification

As mentioned earlier in Chapter 5, when relays are deployed in the network, different handover types appear. These handovers are classified as four main types and object bitrates after any of these handover types can behave differently. Type one is defined as a handover from one relay to another relay. Type two is known as a handover form one relay to an eNB. Type three is considered as a handover from an eNB to a relay and type four is distinguished for the handover between eNBs. The reason for doing this part of study is to show how each handover type effects on object bitrate.

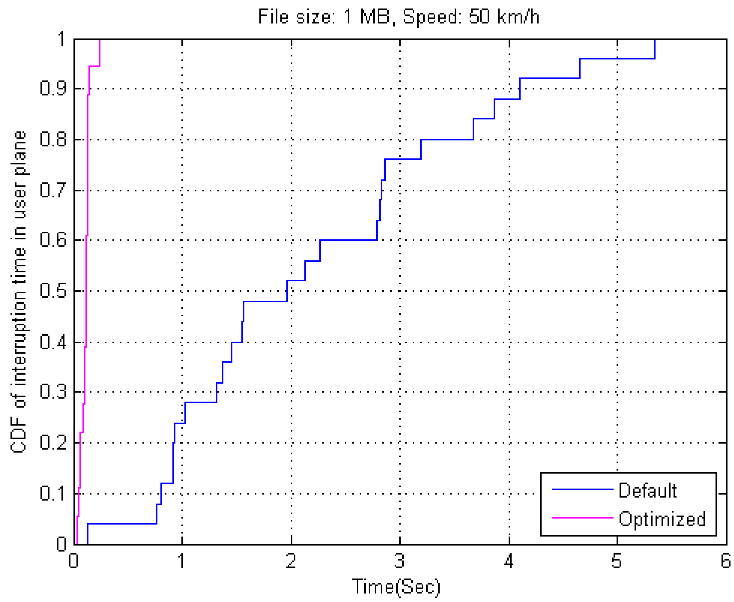

Figure 23: CDF for interruption time in user plane with default and optimized approaches, file size: 1 MB, speed: 50 km/h.

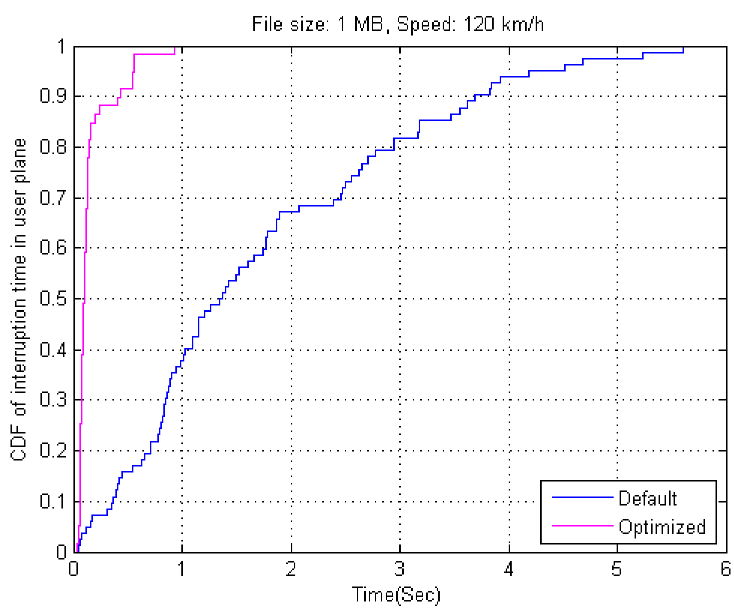

Figure 24: CDF for interruption time in user plane with default and optimized approaches, file size: 1 MB, speed: 120 km/h.

The users under relays in general receive smaller maximum bitrate compared to the users connected directly to the eNBs. This is because relays do not have as many resources as eNBs do. Thus the UE under a relay can never enjoy the maximum bitrate that it can receive under an eNB. Also we have mentioned that the back and forth forwarding problem happens only when the UE is initially connected to

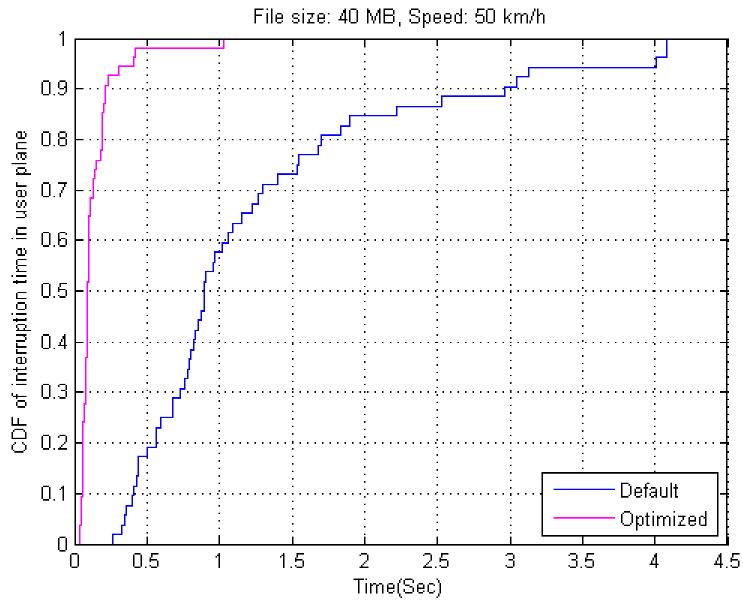

Figure 25: CDF for interruption time in user plane with default and optimized approaches, file size: 40 MB, speed: 50 km/h.

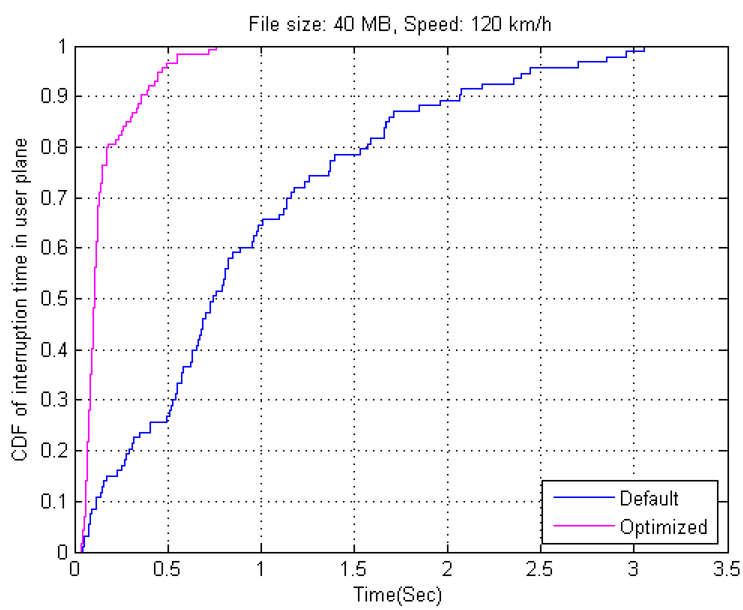

Figure 26: CDF for interruption time in user plane with default and optimized approaches, file size: 40 MB, speed: 120 km/h.

the relay. Thus two types of handovers, RN-RN and RN-DeNB have the back and forth forwarding problem, which produces longer delays. For the RN-RN handover, both connection to the RN and delay of back and forth forwarding problem are considered to have a negative affect on object bitrate. While in the handover from the RN to the DeNB, connection to the DeNB may help the bitrate to increase but delay in data forwarding is again considered as a drawback. For the handover from the DeNB to the RN, there is no back and forth forwarding problem, but because the destination is an RN, the bitrate may decrease.

To conclude, for the RN-RN handovers, we expect the lowest bitrate, for the eNB to the eNB we expect the highest bitrate and for the other two handover types the bitrate ranking might vary. Figure 27 shows the object bitrate after different handover types. In this figure the results of evaluation demonstrate that the RN-RN handover shows the worst bitrate, and eNB-eNB shows the best performance. In the optimized approach, the delay is very small and the dominant effect is due to connection to the DeNB. Therefore in the RN-DeNB handovers, bitrates are better than in the DeNB-RN handovers.

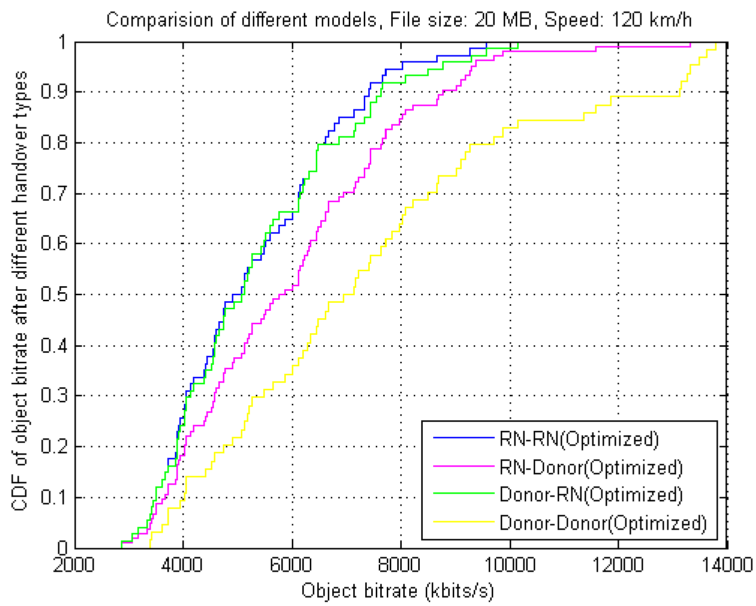

Figure 27: CDF for object bitrate after different handover types, file size: 20 MB, speed: 120 km/h.

6.3.8 Object bitrate comparison

So far we have shown how different approaches behave regarding the amount of forwarded data and delay. We have also shown the effect of different handover types on user bitrate. Now we study how different approaches affect the received object bitrate and how much the optimized approach can improve it. The object bitrate is defined as the file size devided by the file transfer delay.

In Figures 28 and 29, the object bitrate of the default and optimized approaches is illustrated. In Figure 28, the file size is 20 MB and the speed is 50 km/h. On average the bitrate of the optimized approach is 26.5% better than the default approach. In

Figure 29, where the file size is 20 MB and the speed is 120 km/h, the bitrate is improved 21.4% as compared to the default approach.

In Figures 30 and 31, the bitrate is considered for the file size of 30 MB and speeds 50 km/h and 120 km/h. In Figure 30, the bitrate increases 6.2% and in Figure 31 when the speed is 120 km/h the average bitrate increases 13.8% when the optimized approach is compared to the default approach.

In Figures 32 and 33, the file size is 40 MB and the speed is considered as 50 km/h and 120km/h. In Figure 32 the average bitrate increases 12.3% when comparing default approach and the optimized approach. In Figure 33, the optimized approach increases the bitrate by 29.2% comparing to the default approach.

In the last experiment the file size of 50 MB is selected with the speeds of 50 km/h and 120 km/h. In Figure 34, the increase in the bitrate is 8% using optimized approach instead of default approach. In Figure 35, the average increase in object bitrate by deploying optimized approach is 21.3%.

Looking into different approaches, the default, the optimized and the ideal approaches performing lossless handovers, but the dropping PDCP SDUs is not performing a lossless handover. The value of the dropping PDCP SDUs approach is to avoid any data transmission in the uplink of Un link while data forwarding. Figures 35 and 36, show this approach can sometimes show better performance than the default approach. In the ideal approach, there is no delay in delivering the packets and thus this approach should provide the best bitrate. In optimized approach, the packets are received by the receiver after some delay due to data forwarding, we expect this approach to show worse bitrate compared to the ideal approach. The default approach has longer delay than optimized approach, therefore the object bitrate of default approach is worse than optimized approach. In the dropping PDCP SDUs the time for packet delivery is shorter than the default approach, but some packets are dropped. Thus the behavior of this approach varies compared to the other approaches.

It should be noted that the UE can make all different handover types, having different effects on bitrates. For instance, if in simulation time the UE is connected to the RN most of the time and makes RN-RN or RN-DeNB handovers more, this will be seen as a lower bitrate in the end.

In Figures 36 and 37, the object bitrate of the default, the optimized, the dropping PDCP PDUs and the ideal approaches are illustrated. It can be seen in the Figure 36 where the file size is 30 MB and the speed is 50 km/h, the optimized approach shows more similar performance to Ideal approach comparing to the other approaches. As mentioned earlier, for the dropping PDCP SDUs the results are variable. It can be seen in Figure 36 that until 70 percentile, the performance of dropping PDCP SDUs is better than the default approach and from 70 percentile, the trend changes and the default approach shows better performance. In Figure 37, where the file size is 40 MB and the speed is 120 km/h, the optimized approach is similar as the ideal approach. Dropping PDCP PDUs shows better performance than the default approach.

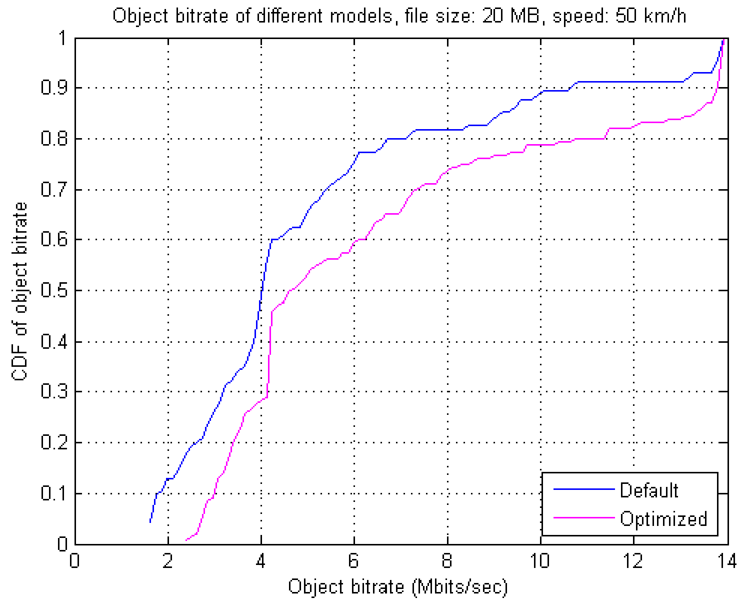

Figure 28: CDF for object bitrate with default and optimized approaches, file size: 20 MB, speed: 50 km/h.

6.3.9 The effect of Active Queue Management (AQM) on bitrate

AQM plays important role to reduce or avoid the probability of severe congestion. AQM does not keep the queues full and gives the sources sufficient time to react to congestion before queues become full. So far all the experiment were done using AQM. In this section the same experiments are performed for different approaches with the identical situations as in 6.4, but without AQM. It is observed in the scenario without usage of AQM, the optimized approach can increase the bitrate up to 46% comparing to the default approach. As it was discussed for the networks using AQM, the bitrate increased up to 30%.

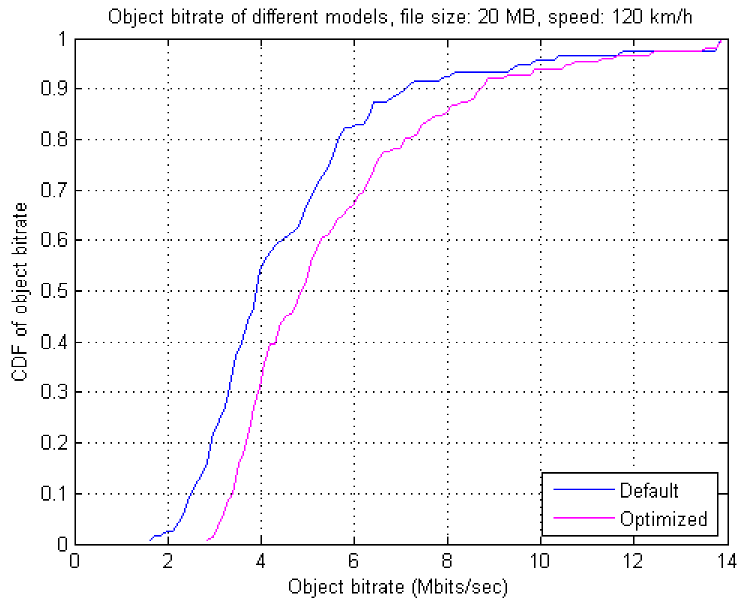

Figure 29: CDF for object bitrate with default and optimized approaches, file size: 20 MB, speed: 120 km/h.

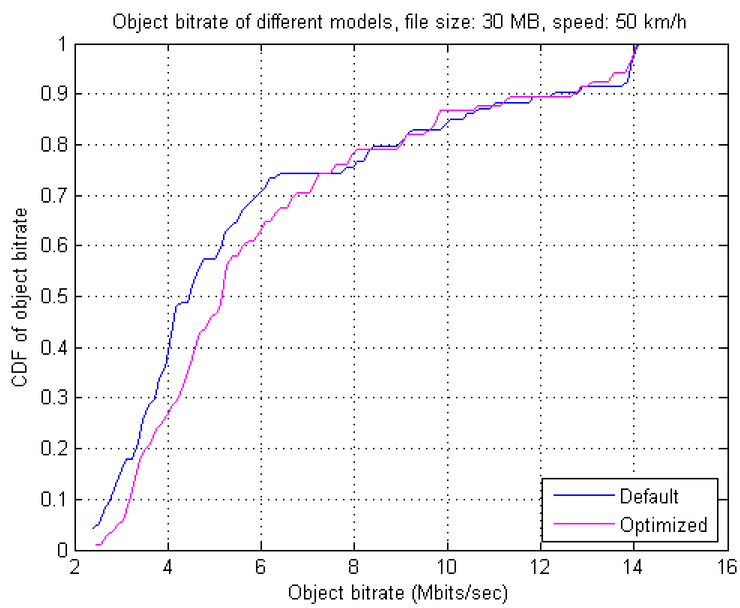

Figure 30: CDF for object bitrate with default and optimized approaches, file size: 30 MB, speed: 50 km/h.

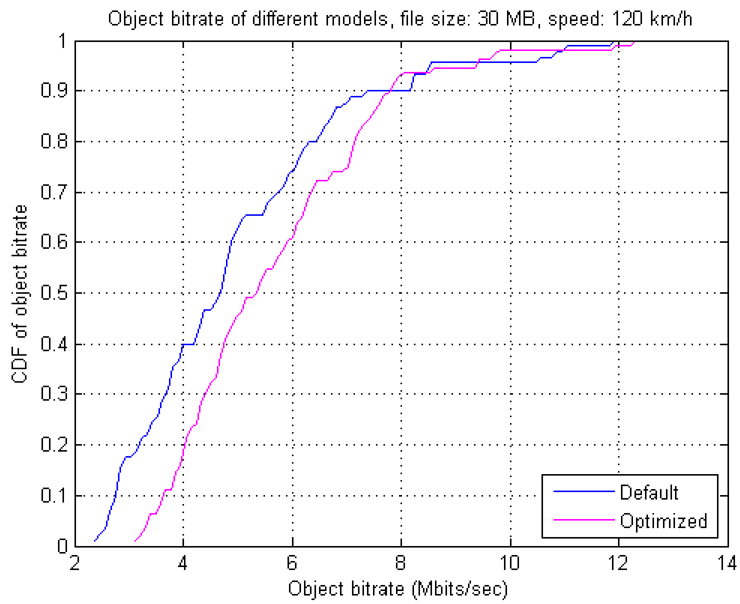

Figure 31: CDF for object bitrate with default and optimized approaches, file size: 30 MB, speed: 120 km/h.

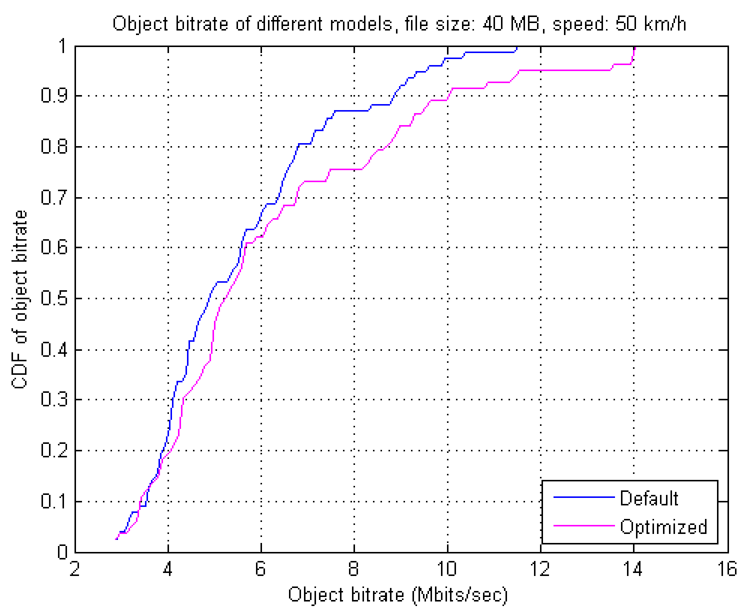

Figure 32: CDF for object bitrate with default and optimized approaches, file size: 40 MB, speed: 50 km/h.

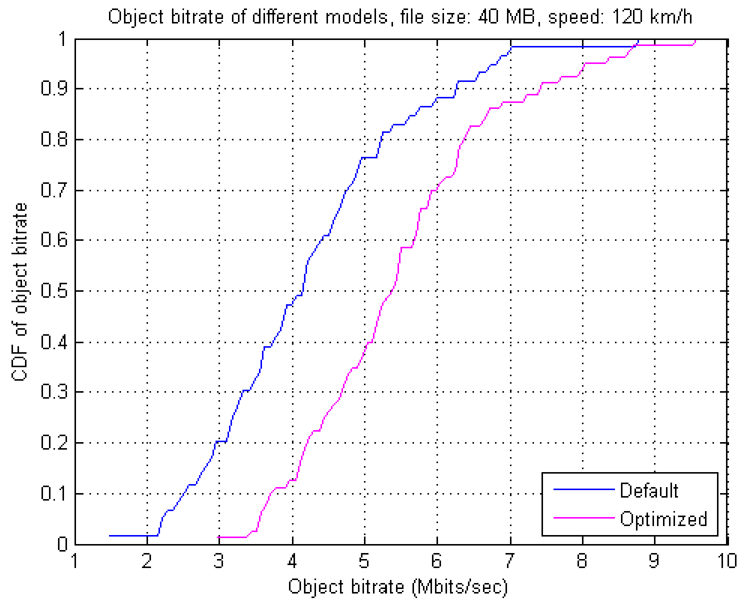

Figure 33: CDF for object bitrate with default and optimized approaches, file size: 40 MB, speed: 120 km/h.

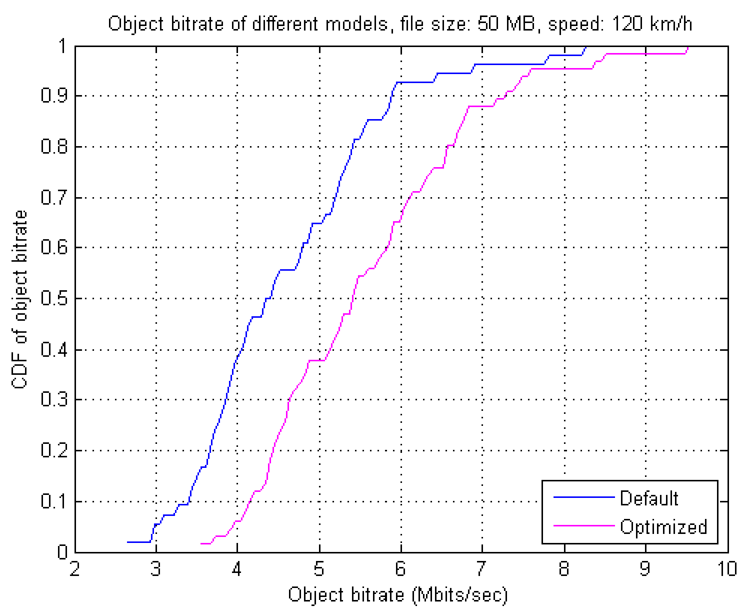

Figure 34: CDF for object bitrate with default and optimized approaches, file size: 50 MB, speed: 50 km/h.

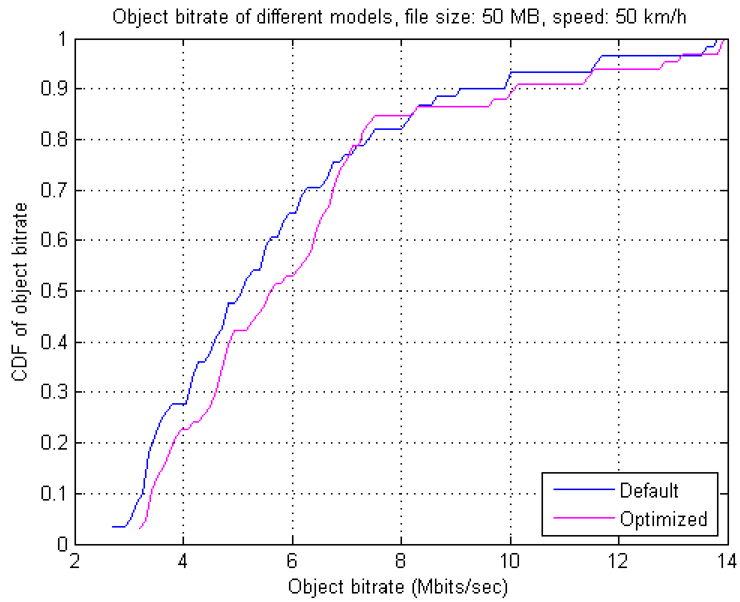

Figure 35: CDF for object bitrate with default and optimized approaches, file size: 50 MB, speed: 120 km/h.

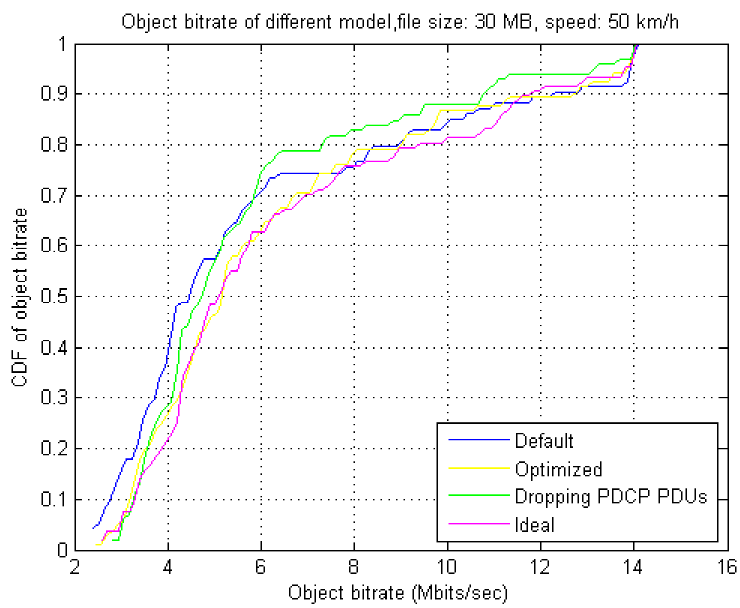

Figure 36: CDF for interruption time in user plane with all four approaches, file size: 30 MB, speed: 50 km/h.

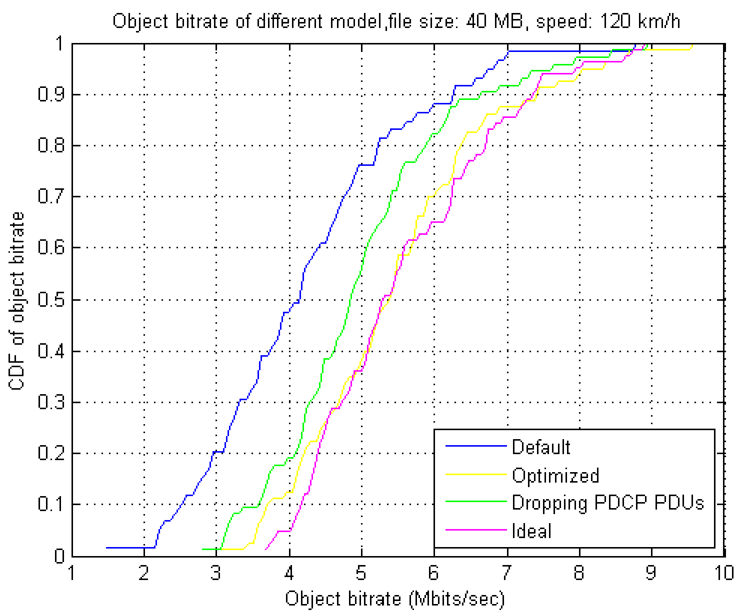

Figure 37: CDF for interruption time in user plane with all four approaches, file size: 40 MB, speed: 120 km/h.

7 Conclusions and future work

7.1 Conclusions

In this thesis, data forwarding in LTE networks containing relays was studied. It was discussed that deploying relays in LTE networks could introduce challenges and it was mentioned that one of these challenges was in handover. In LTE networks containing relays, when the UE which is initially connected to the relay is handed over to the target cell, some amount of data is transmitted back and forth between the relay and the DeNB in data forwarding process. Back and forth data transmission increases the delay and occupies Un resources. This challenge was defined in detail in Chapter 5 and several solutions were addressed to deal with the problem and the advantages and drawbacks of those solutions were discussed.

In Chapter 6, in order to evaluate the proposed solutions in the literature and make comparison, different approaches were developed and simulated. These approaches were named as default, optimized, dropping PDCP SDUs and ideal. The default approach represents the data forwarding approach in LTE networks. The optimized approach tries to make the handover more efficient. Dropping PDCP SDUs does not provide lossless handover, but because it can completely avoid transmission of any data through the Un link during data forwarding time, it has its own benefits to be proposed. The ideal approach is a lossless approach in which packets are delivered without any delay. All the approaches were simulated in a simulator using a variety of scenarios. Different metrics were applied in order to have some clear idea of performance. Metrics such as the amount of bits transferred in handover time, delay, and object bitrate as the ultimate goal in the network were measured and discussed.

Simulations were performed in different speeds and for different file sizes. The selected file sizes were 1 MB, 20 MB, 30 MB, 40 MB and 50 MB. The selected speeds were 50 km/h and 120 km/h.

In the first part of the study the behavior of default and optimized approaches regarding the amount of forwarded data through the Un link was considered. It was seen that applying the optimized approach can decrease the amount of forwarded data passing through the Un link by at least 80%. The optimized approach redirects the data from the source to the destination directly. Therefore the data will reach the destination without being transmitted back and forth in the Un link.

Transmitting forwarded data in the Un link back and forth introduces delay. Especially transmission in the uplink is a limiting factor on data forwarding speed. The delay has a direct relation to the amount of forwarded data. As the optimized approach could decrease the amount of forwarded data passing through the Un link, it could decrease the interruption time in the user plane. The simulations showed that with the use of the optimized approach, the interruption time in the user plane could be reduced by more than 80%.

Deploying relays in LTE networks, different handover types regarding the source and destination of handover appears. In the studies these handover types were categorized into four major groups; RN-RN, RN-eNB, eNB-RN and eNB-eNB. The object bitrate for these handovers were calculated and these measurements were categorized based on handover types. It was seen that, for RN-RN handovers, the object bitrate was the worst while for eNB-eNB handovers the object bitrate was the best. For the other two types the performance varied.

In all kinds of handover scenarios the goal of data forwarding optimization is to increase the performance of the UE, thus the object bitrate using different approaches was simulated and compared. It was seen that using the optimized approach more than 30% better object bitrate was obtained. In the simulations having different file sizes and different speeds, no specific trend in performance was observed.

The effect of AQM was also investigated. It was seen that in the scenarios without AQM, the optimized approach could increase the UE object bitrate by up to 46% compared to the default approach.

7.2 Future work

More work can be done in this interesting area.

1. It would be very useful to investigate multi-hop networks. In that scenario, one RN is served by another RN and that RN again serves another RN. Also sudying data forwarding in cooperative distributive networks would be interesting.

2. It can also be beneficial to study the performance of data forwarding in terms of delay and Un resource consumption in loaded networks.

References

[1] Akan, A. Edemen, C. *Path to 4G wireless networks.* IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications Workshops (PIMRC Workshops), Pages 405–407, 2010.

[2] Abeta, S. *Toward LTE commercial launch and future plan for LTE enhancements (LTE-Advanced).* IEEE International Conference on Communication Systems (ICCS), Pages 146–150, 2010.

[3] Hiramatsu, K., Nakao, S., Hoshino, M., Imamura, D. *Technology evolutions in LTE/LTE-advanced and its applications.* IEEE International Conference on Communication Systems (ICCS), Pages 161–165, 2010.

[4] 3GPP TS 36.101, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); *User Equipment (UE) radio transmission and reception.* V10.2.1, 2011.

[5] Stefania Sesia, Issam Toufik, Matthew Baker, *LTE-The UMTS long Term Evolution: From Theory to Practice.* John Wiley & Sons Ltd, 2009.

[6] Harri Holma, Antti Toskala,*LTE-The UMTS long Term Evolution: From Theory to Practice.* John Wiley & Sons Ltd, 2009.

[7] 3GPP TS 36.323, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); *Packet Data Convergence Protocol (PDCP) specification.* V10.0.0, 2010.

[8] 3GPP TS 36.300, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN);*Overall description.* V10.2.0, 2010.

[9] Jihai Han, Bingyang Wu, *Handover in the 3GPP long term evolution (LTE) systems.* IEEE Global Mobile Congress (GMC), Pages 1–6, 2010.

[10] Racz, A., Temesvary, A., Reider, N. *Handover Performance in 3GPP Long Term Evolution (LTE) Systems.* IEEE Mobile and Wireless Communications Summit,16th IST, Pages 1-5, 2007.

[11] Konstantinos Dimou, Min Wang, Yu Yang, Muhammmad Kazmi, Anna Larmo, Jonas Pettersson, Walter Muller, Ylva Timner, Reider, N. *Handover within 3GPP LTE: Design Principles and Performance.* IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), Pages 1–5, 2009.

[12] Qureshi, S.-U.-R., Nawaz, S.J., Patwary, M., Abdel-Maguid, M., Kamar, A., Reider, N. *The impact of propagation environment and velocity on the handover performance of LTE systems.* IEEE International Conference on Wireless Communications and Signal Processing (WCSP), Pages 1–5, 2010.

[13] Huaining Ge, Xiangming Wen, Wei Zheng, Zhaoming Lu, Bo Wang, *A History-Based Handover Prediction for LTE Systems.* Computer Network and Multimedia Technology, IEEE International Symposium on, Pages 1–4, 2009.

[14] Tae-Hyong Kim, Qiping Yang, Jae-Hyoung Lee, Soon-Gi Park, Yeon-Seung Shin *A Mobility Management Technique with Simple Handover Prediction for 3G LTE Systems.* IEEE 66th Vehicular Technology Conference, Pages 259–263, 2007.

[15] Legg, P., Gao Hui, Johansson, J. *A Simulation Study of LTE Intra-Frequency Handover Performance.* IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall), Pages 1–5, 2010.

[16] Yejee Lee, Bongjhin Shin, Jaechan Lim, Daehyoung Hong, *Effects of time-to-trigger parameter on handover performance in SON-based LTE systems.* IEEE 16th Asia-Pacific Conference on Communications (APCC), Pages 492–496, 2010.

[17] Anas, M., Calabrese, F.D. Mogensen, P.E. Rosa, C. Pedersen, K.I. *Performance Evaluation of Received Signal Strength Based Hard Handover for UTRAN LTE.* IEEE 65th Vehicular Technology Conference, Pages 1046–1050 ,2010.

[18] Singh, B. *Hard Handover Performance Evaluation through Link Drops.* Signal Processing, IEEE International Conference on Communications and Networking, pages 459–463, 2007.

[19] Kwan, R., Arnott, R. Paterson, R. Trivisonno, R. Kubota, M. *On Mobility Load Balancing for LTE Systems.* IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall),Pages 1–5, 2010

[20] Tarjan, P., Nemeth, G. *Buffer Overflow Probability of TCP Flows During Mobile Handovers.* IEEE Communications Letters, Volume 12, Pages 481–483, 2008.

[21] Lan Wang, Yongsheng Zhang, Zhenrong Wei, *Mobility Management Schemes at Radio Network Layer for LTE Femtocells.* IEEE 69th Vehicular Technology Conference, Pages 1–5, 2009.

[22] Naizheng Zheng Wigard, J. *On the Performance of Integrator Handover Algorithm in LTE Networks.* IEEE 68th Vehicular Technology Conference, Pages 1–5, 2008.

[23] Wang Weidong Zhang Yinghai Zhang Heng Zhou Yao *Sub-carriers bidirectional arrayed handover algorithm.* IEEE 2nd International Conference on Broadband Network & Multimedia Technology, Pages 75–79, 2009.

[24] Bajzik, L. Horvath, P. Korossy, L. Vulkan, C. *Impact of Intra-LTE Handover with Forwarding on the User Connections.* IEEE 16th Mobile and Wireless Communications Summit, Pages 1 - 5, 2007.

[25] Pacifico, D. Pacifico, M. Fischione, C. Hjalrmasson, H. Johansson, K.H. *Improving TCP Performance During the Intra LTE Handover.* IEEE Global Telecommunications Conference, Pages 1–8, 2009.

[26] Kitatsuji, Y. Yokota, H. *On Handover Procedure with Data Forwarding for Reducing Buffered User Data in Base Stations.* IEEE Global Telecommunications Conference, Pages 1–8, 2009.

[27] Usui, T., Kitatsuji, Y., Yokota, H. *On reducing packet buffering for network-based fast handover cooperating with IMS.* IEEE International Conference on Internet Multimedia Services Architecture and Applications (IMSAA), Pages 1–6, 2009.

[28] Gohar, M. Seok Joo Koh Tae-Won Um Hyun-Woo Lee, *Seamless multicast handover in PMIPv6-based wireless networks.* IEEE 12th International Conference on Advanced Communication Technology (ICACT), 2010 The 12th International Conference on, Volume 1, Pages 502–507, 2010.

[29] Saito, M. Morimoto, A. Takyu, O. Nakagawa, M. *IP Packet Loss Compensation Scheme with Bicast and Forwarding for Handover in Mobile Communications.* IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, Page(s): 1–5, 2006.

[30] Seong-Soo Park, Young-Jae Song, Dong-Ho Cho, *Performance improvements of mobile data protocol in wire and wireless interworking environments.* IEEE Global Telecommunications Conference, Volume 5, Pages 2646 - 2651, 1998.

[31] Widjaja, Indra, La Roche, Humberto, *Sizing X2 Bandwidth for Inter-Connected eNBs.* IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)Page(s): Pages 1–5, 2009.

[32] Thomas Wirth, Venkatkumar, V., Thomas Haustein, Egon Schulz, Rudiger Halfmann, *LTE-Advanced Relaying for Outdoor Range Extension.* IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), Pages: 1 - 4, 2009.

[33] Jing Han, Haiming Wang, *Uplink Performance Evaluation of Wireless Self-Backhauling Relay in LTE-Advanced.* IEEE 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), Pages 1–4, 2010.

[34] Abdallah Bou Saleh, Simone Redana, Bernhard Raaf, Jyri Hamalainen, *Comparison of Relay and Pico eNB Deployments in LTE-Advanced.* IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), Pages 1–5, 2009.

[35] Beniero, T. Redana, S. Hamalainen, J. Raaf, B. *Effect of Relaying on Coverage in 3GPP LTE-Advanced.* IEEE 69th Vehicular Technology Conference,Pages 1–5, 2009.

[36] Wirth, T. Thiele, L. Haustein, T. Braz, O. Stefanik, J. *LTE Amplify and Forward Relaying for Indoor Coverage Extension.* IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall), Pages 1–5, 2010.

[37] Zhangchao Ma, Yi Zhang, Kan Zheng, Wenbo Wang, Muzi Wu, *Performance of 3GPP LTE-Advanced networks with Type I relay nodes.* 5th International ICST Conference on Communications and Networking in China (CHINACOM), Pages 1–5, 2010.

[38] Letian Rong, Elayoubi,S.E., Haddada,O.B. *Impact of Relays on LTE-Advanced Performance.* IEEE International Conference on Communications (ICC), Pages 1–6, 2010.

[39] Jinhua Jiang, Wei Heng, Hua Zhang, Gang Wu, Haifeng Wang, *A novel handover scheme for multi-hop transparent relay network.* International Conference on Wireless Communications & Signal Processing, Pages 1–5, 2009.

[40] Jantarayat, I., Uthansakul, P., Uthansakul, M. *Tradeoffs between handover performance and coverage range of relay stations in multihop cellular networks.* International Conference on Wireless Communications and Signal Processing, Pages 1–5, 2010.

[41] Wei Wei, *The fast handover method of Relay Station based on priority message.* International Conference on Computer Application and System Modeling (ICCASM), Volume 2, Pages 694–696, 2010.

[42] JuYeop Kim, Dong-Ho Cho, *Pre-Buffering Scheme for Seamless Relay Handover in Relay Based Cellular Systems.* IEEE 69th Vehicular Technology Conference, Pages 1–5, 2009.

[43] Motorola, *Handovers involving Type-1 Relay Node.* 3GPP TSG-RAN-WG2 Meeting #66, R2-093207, San Francisco, US, May 4–8, 2009.

[44] Research In Motion UK Limited, *Joint PDCP protocols on Uu and Un interfaces to improve type-I relay handover.* 3GPP TSG RAN WG2 Meeting #66bis, R2-093735, Los Angeles, USA, June 29–July 3, 2009.

[45] LG Electronics Inc, *Consideration on Data Forwarding for Relay based Handover Procedure.* 3GPP TSG-RAN WG3 Meeting #68, R3-101447, Montreal, Canada, May 10–14, 2010.

[46] CATT, *Data forwarding mechanism for supporting RN.* 3GPP TSG-RAN WG3 Meeting #68, R3-101455 Montreal, Canada, May 10–14, 2010.

[47] Alcatel-Lucent, Alcatel-Lucent Shanghai Bell, *Solution for UE handover issue for Relay.* 3GPP TSG-RAN WG3 #68, R3-101609, Montreal, Canada, May 10–14, 2010.

[48] CATT, *Discussion on handover performance.* 3GPP TSG RAN WG2 Meeting #68bis, R2-100177, Valencia, Spain, January 18th–22nd, 2010.

[49] Samsung, *Handover performance of relay architectures.* 3GPP TSG RAN WG2 #67, R2-094879, Shenzhen, China, August 24–28 , 2010.